creative commons

COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

공학전문석사 학위 연구보고서

머신러닝을 이용한 상수도 관로의 노후도 예측 연구

A Study on Prediction of Aging Evaluation Grade of Water Supply Pipeline Using Machine-Learning

2023년 2월

서울대학교 공학전문대학원

응용공학과 응용공학전공

이 종 호

머신러닝을 이용한 상수도 관로의 노후도 예측 연구

A Study on Prediction of Aging Evaluation Grade of Water Supply Pipeline Using Machine-Learning

지도교수 지석호 · 구윤모

이 프로젝트 연구보고서를 공학전문석사
학위 연구보고서로 제출함
2023년 2월

서울대학교 공학전문대학원
응용공학과 응용공학전공
이종호

이종호의 공학전문석사 학위 연구보고서를 인준함
2023년 2월

위 원 장 한 훈 (인)

위 원 지석호 (인)

위 원 구윤모 (인)

초 록

2019년 기준 우리나라의 상수도 보급률은 전국 평균 99.3%로 전국 대부분의 지역에서 상수도 이용이 가능하다. 반면, 20년 이상 경과된 관로의 수는 점차 증가하여 전체 관로의 34.8%인 7.94만km이며, 시간이 경과할수록 계속 증가할 전망이다.

상수도 관로의 노후화는 관로 파손으로 인한 단수, 상수도 내부의 스케일이나 부식 문제로 인한 붉은 수돗물 등 수질문제를 야기하고, 이는 국민의 수돗물에 대한 불신을 증가시킬 수 있다.

따라서, 적기에 노후 상수도 관로를 개량하여 물 공급의 신뢰성을 확보하는 노력이 필요하다. 노후 상수도 관로 개량을 위해 반드시 필요한 절차는 노후도 평가이다. 노후도 평가는 기존의 운영인자를 바탕으로 관로의 노후상태를 간접적으로 측정하는 간접평가와 실제 단수를 통해 관로를 절단하거나 굴착하여 관 내·외부의 부식정도를 직접 측정하는 직접평가로 구분할 수 있다.

하지만, 현행 간접평가와 직접평가 방법은 데이터의 대표성 문제, 평가등급의 차이, 조사과정에서 위험, 단수 리스크 등 한계점이 있다.

이에, 본 연구는 노후관 조사시 간접평가와 직접평가의 한계점을 극복하고, 데이터 기반의 학습방식인 머신러닝을 활용하여 직접조사 없이 실제 직접조사 결과에 근접한 노후도 평가 결과를 예측하기 위해 진행하였다. 그 결과, 92.83%의 정확도를 갖는 노후도 예측 모델을 개발하였다. 향후 실제 사업에 적용하여 노후관 개량사업의 효과성을 제고하고, 단수나 안전문제를 해소할 수 있을 것으로 기대한다.

주요어 : 노후도 평가, 간접평가, 직접평가, 머신러닝, 노후도 예측
학 번 : 2021-20316

목 차

제 1 장 서론 ····· 1
1.1 연구배경 ····· 1
1.2 연구목적 ····· 5
1.3 연구범위 ····· 6
1.4 연구방법 ····· 8

제 2 장 기존 노후도 평가방법과 한계점 ····· 9
2.1 노후도 평가 관련 선행연구 고찰 ····· 9
2.2 상수도 강관의 특성과 부식 매커니즘 ····· 11
2.3 상수도 강관의 노후도 평가방법 ····· 16
2.4 상수도 강관의 간접평가 방법 ····· 18
2.5 상수도 강관의 직접평가 방법 ····· 20
2.6 기존 상수도 강관의 노후도 평가 한계점 ····· 23
2.7 기존 노후도 평가 한계점 개선을 위한 연구방향 ·· 32

제 3 장 데이터 특성분석을 통한 변수 최적화 ·· 33
3.1 노후도 평가 데이터 수집 ····· 33
3.2 노후도 평가 데이터의 특성 분석 ····· 35
3.3 데이터의 특성에 따른 독립변수 채택 ····· 49
3.3 노후도 예측모델의 종속변수 채택 ····· 51

제 4 장 강관의 노후도 예측 모델 개발 ····· 52
4.1 분석방법의 결정 ····· 52
4.2 분석 알고리즘 및 성능평가 방법 ····· 54
4.3 독립변수 채택의 적정성 검증 ····· 57
4.4 노후관 개량사업 대상여부 예측 모델 ····· 60
4.5 직접평가 등급 예측 모델 ····· 67
4.6 Class Ⅲ과 Class Ⅳ의 이진분류 모델 ····· 74
4.7 실제 노후관 개량사업 적용 결과 ····· 82

제 5 장 결론 ····· 83
5.1 연구결과 ····· 83
5.2 연구의 한계점 ····· 87
5.3 연구 성과 ····· 89
5.4 향후 연구 ····· 90

참고문헌 ····· 91
Abstract ····· 93

표 목 차

[표 1] 경과년수별 수도관 연장 현황 ····· 1
[표 2] 상수도 관로의 노후도 평가 단계 ····· 17
[표 3] 상수도 강관의 간접평가 항목 ····· 18
[표 4] 상수도 강관의 간접평가 점수 산정방법 ····· 19
[표 5] 간접평가와 직접평가 등급 비교표 ····· 19
[표 6] 상수도 강관의 직접평가 항목 ····· 20
[표 7] 직접평가 등급에 따른 관의 상태와 개량방법 ····· 22
[표 8] 분석데이터의 간접, 직접평가 분포표 ····· 24
[표 9] 관경별, 노후관 등급별 단위 공사비 ····· 27
[표 10] 관경별 단수 시 예상 단수인구 현황 ····· 31
[표 11] 내부 압력에 의한 원주방향응력 산정공식 ····· 35
[표 12] 직접평가 등급별 평균 직경 현황 ····· 36
[표 13] LI 수질부식성 지표 산정 공식 ····· 39
[표 14] 데이터 특성을 고려한 독립변수 채택 사유 ····· 50
[표 15] 직접평가 등급별 데이터의 수와 개량방법 ····· 51
[표 16] 분류모델의 성능검증을 위한 오분류표 ····· 56
[표 17] Case별 노후관 개량대상여부 예측 성능 비교 ····· 57
[표 18] Case별 직접평가 등급 예측 성능 비교 ····· 58
[표 19] 분류 알고리즘별 노후관 개량사업 대상 예측 성능 ····· 66
[표 20] 분류 알고리즘별 직접평가 등급 예측 성능 ····· 72
[표 21] 2가지 알고리즘의 Class별 Sensitivity ····· 73
[표 22] 독립변수 재검토 결과 및 변수 채택사유 ····· 81
[표 23] 2가지 알고리즘별 Ⅲ, Ⅳ등급 분류 성능 ····· 76
[표 24] 실제 노후관 개량사업 적용 결과 오분류표 ····· 82
[표 25] 정확도 개선에 따른 정량적 기대효과 ····· 84
[표 26] 예산절감액 산출근거 ····· 89

그 림 목 차

[그림 1] 광역상수도 공급량과 관종별 현황 ····· 6
[그림 2] 관경별 경제적 유량 ····· 7
[그림 3] 상수도용 강관의 구조 ····· 11
[그림 4] 매설년수별 내부 도장재의 박리비율 ····· 12
[그림 5] 상수도 강관의 부식 매커니즘 ····· 13
[그림 6] 상수도 강관의 파손 진행단계 ····· 15
[그림 7] 강관 직접평가 등급 판정 흐름도 ····· 21
[그림 8] 관 노후도 조사 전경 ····· 28
[그림 9] 일반성 재해와 질식재해의 사망률 ····· 29
[그림 10] 안전사고 예방을 위한 효과적 관리방안 ····· 30
[그림 11] 연구방향(AS-IS TO-BE) ····· 32
[그림 12] 분석데이터 현황 ····· 33
[그림 13] 분석데이터의 기초 통계량 ····· 34
[그림 14] 직접평가 등급과 직경간의 Density Plot ····· 36
[그림 15] 부식두께 예측모델 ····· 37
[그림 16] 직접평가 등급과 매설년도간의 Density Plot ···· 38
[그림 17] 직접평가 등급과 수질부식성간의 Density Plot ··· 39
[그림 18] 직접평가 등급과 토양부식성간의 Density Plot ··· 41
[그림 19] 직접평가 등급과 토양부식성간의 Box Plot ····· 41
[그림 20] 직접평가 등급과 매설깊이간의 Density Plot ····· 44
[그림 21] 직접평가 등급과 매설깊이간의 Box Plot ····· 44
[그림 22] 직접평가 등급과 도로형태 산점도 ····· 45
[그림 23] 직접평가 등급과 내부하중간의 Density Plot ····· 46
[그림 24] 광역상수도 사고 원인별 현황('15~'20) ····· 47

[그림 25] 독립변수의 중요도 ····· 49
[그림 26] 변수간 상관관계 행렬 및 다중공선성 ····· 53
[그림 27] 회귀분석 결과 ····· 54
[그림 28] Case 1의 Overall Statistics ····· 59
[그림 29] Case 2의 Overall Statistics ····· 59
[그림 30] Confusion Matrix(RF_사업대상여부) ····· 60
[그림 31] Confusion Matrix2(RF_사업대상여부) ····· 61
[그림 32] ROC (RF_사업대상여부) ····· 61
[그림 33] Confusion Matrix(SVM_사업대상여부) ····· 62
[그림 34] Confusion Matrix2(SVM_사업대상여부) ····· 63
[그림 35] ROC (SVM_사업대상여부) ····· 63
[그림 36] Confusion Matrix(XGBoost_사업대상여부) ··· 64
[그림 37] Confusion Matrix2(XGBoost_사업대상여부) · 65
[그림 38] ROC (XGBoost_사업대상여부) ····· 65
[그림 39] OOB error rate (RF_직접평가 등급) ····· 67
[그림 40] Confusion Matrix(RF_직접평가 등급) ····· 68
[그림 41] Confusion Matrix2(RF_직접평가 등급) ····· 68
[그림 42] Confusion Matrix(SVM_직접평가 등급) ····· 69
[그림 43] Confusion Matrix2(SVM_직접평가 등급) ····· 70
[그림 44] ROC (SVM_직접평가 등급) ····· 70
[그림 45] Confusion Matrix(XGBoost_직접평가 등급) ···· 71
[그림 46] Confusion Matrix2(XGBoost_직접평가 등급) ··· 72
[그림 47] 독립변수 중요도 (Ⅲ, Ⅳ등급 이진분류) ····· 75
[그림 48] 강관의 안전계수 계산 흐름도 ····· 76
[그림 49] Confusion Matrix(RF_직접평가 Ⅲ, Ⅳ등급) ····· 77
[그림 50] Confusion Matrix2(RF_직접평가 Ⅲ, Ⅳ등급) ··· 78
[그림 51] ROC (RF_사업대상여부) ····· 78

[그림 52] Confusion Matrix(XGBoost_직접평가 Ⅲ, Ⅳ등급) ···· 79
[그림 53] Confusion Matrix2(XGBoost_직접평가 Ⅲ, Ⅳ등급) ··· 80
[그림 54] ROC (XGBoost_직접평가 Ⅲ, Ⅳ등급) ····· 80
[그림 55] 외부 하중에 의한 휨응력 Density plot ····· 86
[그림 56] 향후 데이터 수집 및 분석체계 ····· 88

제1장 서론

1.1 연구배경

2020년 상수도통계(환경부)에 따르면 전국 총인구 52,975천명의 99.4%인 52,644천명에게 상수도를 공급하고 있다. 이제는 산간벽지를 제외한 전국 어느 곳에서나 상수도 혜택을 누릴 수 있다. 상수도를 공급하는 전국 수도관 228,223km 중 34.8%인 79,461km는 설치 후 21년 이상 경과된 경년관에 해당된다. [1]

[표 1] 경과년수별 수도관 연장 현황

(단위 : km)

구 분	총 계	5년 이내	6~10년	11~15년	16~20년	21년 이상
총 계	228,323 (100%)	35,860 (15.7%)	41,921 (18.4%)	39,684 (17.4%)	31,397 (13.7%)	**79,461** **(34.8%)**
지방상수도	222,641 (100%)	35,487 (15.9%)	41,523 (18.7%)	39,043 (17.5%)	29,758 (13.4%)	**76,830** **(34.5%)**
광역상수도	5,682 (100%)	373 (6.6%)	399 (7.0%)	640 (11.3%)	1,639 (28.8%)	**2,631** **(46.3%)**

상하수도 협회 용어해설에 따르면 경년관(經年管, aged pipes)이란 포설(布設) 후 오랜 기간이 경과한 수도관으로서 일반적으로 사용연수의 경과와 함께 노후화되어 누수사고 발생의 위험도가 높아지고, 적수발생[1]이나 출수불량[2]의 원인이 되는 관이라고 정의한다. 하지만,

1) 적수발생 : 매설 후 오랜 시간 사용한 상수도관 내부의 압력변화나 충격 등에 의해 관 내부의 녹이나 이물질이 수돗물에 포함되어 수돗물이 붉게 보이는 현상
2) 출수불량 : 관 내부에 스케일이 쌓여 관로의 통수단면적을 축소시키고, 조도를 불량하게 하여 손실을 증가하면서 고지대 등에 물이 약하게 나오는 현상

노후화의 진행은 관의 재질, 관 내부 유체의 수질, 유속, 매설 환경에 따라 달라 21년이상 경년관이 반드시 관로를 새로 바꾸거나 개량할 필요가 있는 '노후관'은 아니라고 정의하고 있다. [2]

「상수도 관망의 기술진단 범위 및 시행방법 등에 관한 고시」(환경부고시 제2007-11호)에 따르면 노후관은 누수가 잦은 관, 구조적 강도가 저하된 관, 관 내부에 녹이 발생하여 녹물이 많이 나오는 관 등 수도관으로서 제 기능을 발휘하지 못하는 관을 말한다. [3]

즉, 매설 후 시간이 경과하였다고 하여 반드시 노후관이라고 정의할 수는 없으며, 구조적 강도 저하여부, 관 내부의 녹 발생여부와 같이 수도관이 제 기능을 하고 있는지에 대한 노후도 평가를 실시하여야 하고, 노후관으로 평가되면 관 교체, 갱생 등 적정한 개량방안을 강구하여야 한다.

상수관로의 노후화와 파손은 매설연수와 매설환경, 운영조건 등 여러 가지 복합적인 요인에 의해 발생한다. 상수관로가 지중에 매설되면 매설환경 및 운영조건에 따라 토압, 차량하중 등에 의해 응력이 발생하고, 관체 내·외부의 부식진행정도에 따라 안전계수가 감소하게 되며, 이론적으로 안전계수가 '1'이하가 되면 파손이 발생한다.[4]

상수도 관로가 파손되면 물 공급 중단이 발생할 수 있고, 노후화로 인해 관 내부에 녹이나 이물질이 많으면 붉은 수돗물(적수)과 같이 오염된 수돗물이 공급될 수 있어 국민의 수돗물 이용에 많은 불편이 발생될 수 있다. 따라서 상수도 관로의 파손과 노후화 심화 이전에 적정한 개량사업 진행하는 것이 필요하며, 이를 위해서 노후도 조사(관 상태평가)를 실시하고 있다.

일반적으로 관 노후도 조사는 관로제원, 운영이력, 매설환경 등을 고려하여 관의 상태를 추정하기 위한 간접평가 방법(예측모델)과 현장에서 굴착 후 관을 노출하거나 관 내부에 진입하여 상수도 관의 상태를 조사하고 평가하는 직접평가 방법이 있다.

간접평가는 일반적으로 관경, 매설년수, 수질부식성, 토양부식성 등 관의 노후도와 관련된 인자들의 상황에 따라 점수를 부여하고, 각 항목에 대한 가중치를 곱하여 점수로 산정하는 점수평가법을 적용한다. 간접평가(점수평가법)는 우리나라에서 가장 많이 사용하는 평가방법으로서 현장조사가 거의 없고, 비용이 저렴한 점, 단수가 불필요한 점, 데이터 취득이 용이하여 평가가 편리한 장점이 있다. 하지만, 입력자료에 의해 정확성이 달라질 수 있고, 장래 개량을 위한 예측력이 결여되는 점, 현재의 상태만 평가가 가능한 점, 상수도 관로의 간접적인 상태를 예측하는 방법이므로 실제 노후관 개량사업의 의사결정에 직접 활용할 수 없는 한계점이 있다. [5]

이에, 정확한 관로의 상태평가를 위해 직접평가를 실시한다. 직접평가는 표본굴착, 관 내부 진입 후 조사 또는 비파괴 조사장비(부단수 내시경 장비)를 이용하여 관의 상태를 직접 확인하고 조사함으로서 객관적인 조사 결과를 확보할 수 있다. 그러나 직접평가를 위해서는 관 매설지점 굴착, 많은 비용과 시간, 조사를 위한 단수가 수반된다. 또한, 굴착이나 관 내부 조사는 협소하고 밀폐된 공간에서 수행되기 때문에 조사자의 안전사고 위험도가 크게 증가한다. 아울러, 물을 공급중인 관로에서는 단수를 하는 것이 매우 어려워 조사 지점이 매우 제한적이기 때문에 조사결과의 대표성이 부족하다는 한계가 있다.[5]

이처럼 상수도 관로의 노후도 평가는 간접평가와 직접평가 방법

모두 근본적인 한계가 있다. 관로의 상태를 정확히 진단해야 상태에 알맞은 적정한 개량을 실시할 수 있으나 기존의 노후도 평가방법 상 한계로 인해 정확한 관 상태 진단에 어려움이 있는 실정이다.

정부는 2022년 7월 120대 국정과제를 발표하고, 국정과제의 하나로 '기후위기에 강한 물 환경과 자연 생태계 조성'을 목표로 노후화된 물 관리시설을 현대화하는 과제를 추진하고 있다. 이의 일환으로 K-water는 노후관 개량 사업을 포함한 수도시설 안정화 전략계획을 수립하였다. 환경부는 2040년까지 약 12조 3천억 원의 막대한 재원을 투입하여 단계적으로 노후 상수도 관을 개량하는 계획을 2040 국가수도기본계획에 반영하였다. [6]

또한, 정부는 2022년 1월 27일 「중대재해처벌법」을 본격 시행하고 사업 또는 사업장에서 일하는 모든 사람의 안전 및 보건을 확보하도록 경영책임자에게 안전보건의무를 부과하였다. 공공기관 경영에서 안전 우선 원칙을 확립하기 위해 안전 중심 경영평가 실시, 안전투자 대폭 확대 등 안전중심 경영구조로 전환을 요구하고 있다.

이처럼 정부 정책에 따라 앞으로 대규모 노후 상수도 관 개량사업이 진행될 예정이며, 조사나 사업 시행과정에서 안전을 확보하는 것은 기본 의무이자 최우선 가치로 고려해야 하는 상황이다.

향후 진행될 대규모 노후 상수도관 개량사업의 핵심요소는 정확한 상수도 관의 노후도 평가이다. 이에, 기존 노후도 평가방법의 한계점을 개선하여 막대한 재원이 투입되는 노후관 개량의 효과성을 제고하고 조사과정에서의 안전을 확보하는 한편, 단수로 인한 국민 불편 문제를 해소하기 위한 노후도 평가 개선 연구가 필요한 실정이다.

1.2 연구목적

본 연구는 노후관 조사시 간접평가와 직접평가의 한계점을 극복하고, 데이터 기반의 학습방식인 머신러닝을 활용하여 직접조사없이 실제 직접조사 결과에 근접한 노후도 평가 결과를 예측하기 위해 진행하였다. 이를 위해 운영 중 비교적 쉽게 취득할 수 있는 간접평가 데이터(10가지 항목)를 독립변수로 하고, 직접평가를 통해 취득하는 노후도 평가 등급을 종속변수로 하는 머신러닝 기반의 상수도 관로 노후도 예측 모델을 개발하였다. 독립변수는 연구대상인 광역상수도 강관의 부식특성과 간접평가 데이터의 특성을 고려하여 현행 10가지 항목을 5가지 항목으로 모델에 맞도록 최적화하였고, 종속변수는 노후관 개량사업 대상여부, 노후도 직접평가 등급 예측 등 2가지 항목으로 하였다.

본 연구 모델을 활용하면 직접조사에 소요되는 시간과 비용, 전문인력의 투입을 획기적으로 줄일 수 있으며, 간접평가 결과만 있으면 조사구간을 얼마든지 세분화할 수 있어 제한적인 직접조사의 한계점도 극복할 수 있다. 특히, 직접조사로 인한 안전사고 위험과 단수 리스크를 제거할 수 있다.

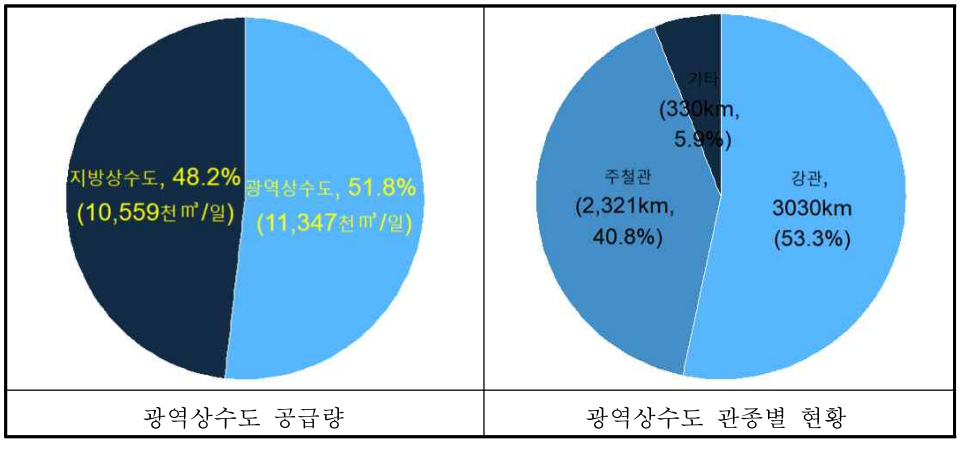
연구모델을 이용한 상수도관 노후도 예측이 실제 사업수행 시 정착되면 조사예산 절감, 기업의 안전경영 기여, 조속한 노후관 개량사업 추진으로 수돗물에 대한 국민의 신뢰도 제고에 기여할 것으로 기대한다.

1.3 연구범위

본 연구는 전국 상수도 공급량의 51.8%를 차지하는 광역상수도[3] 시설 중 강관(Steel Pipe)의 노후도를 예측하는 것을 연구 범위로 한다.

2020년 수도시설 운영관리 편람(K-water)에 따르면 전국 상수도 공급량 중 51.8%인 11,347천㎥/일은 광역상수도를 통해 공급하고 있다. 이를 1인 1일당 급수량(LPCD, 345ℓ/인・일)으로 환산하면 전국 32,890천명이 광역상수도를 이용하여 물을 사용하고 있다. 광역상수도는 댐이나 하천 등 취수원에서 취수한 원수 또는 정수(정수장에서 처리한 수돗물)를 지방자치단체에 공급하고, 지방자치단체는 자체 수도시설을 이용하여 각 가정집과 산업단지 등 수용가에 공급한다.[7]

따라서, 광역상수도 시설 노후화로 인해 수질, 수량, 수압 상의 문제가 발생할 경우 수많은 국민이 수돗물 이용에 불편을 겪게 된다. 이에, 전국 상수도 시설의 51.8%를 공급하는 광역상수도 시설을 연구 범위로 결정하였다.



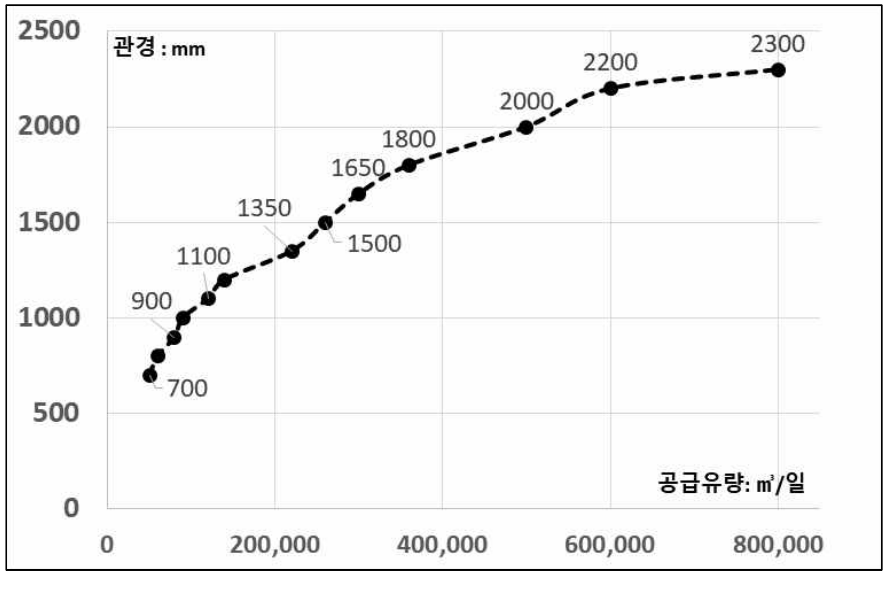
광역상수도 공급량	광역상수도 관종별 현황

[그림 1] 광역상수도 공급량과 관종별 현황

3) 수도법 제3조(정의)에 따라 국가, 지방자치단체, 한국수자원공사 또는 환경부 장관이 인정하는 자가 둘 이상의 지방자체단체에 원수나 정수를 공급하는 일반수도

연구 범위 결정 시 관의 종류는 강관으로 결정하였다.

광역상수도 관로 총 연장 5,681km 중 관종별 구성비율은 강관이 53.3%, 주철관 40.8%, 기타 5.9%로 구성되어 있다. 이에, 광역상수도 관로시설 중 강관이 가장 많은 점, 강관은 대부분 직경 1,000mm 이상 대형관에서 사용하는데 관의 직경이 클수록 물 공급량이 많고 문제 발생시 파급영향이 매우 큰 점, 그간 진행되었던 노후관 개량사업의 대부분은 강관에서 시행된 점, 본 연구는 데이터 기반 학습으로 수행하였기 때문에 선행 사업에서 노후도 데이터를 취득해야 하는 점을 종합적으로 고려하여 강관을 연구범위로 결정하였다.



[그림 2] 관경별 경제적 유량

1.4 연구방법

본 연구는 머신러닝(Machine Learning) 기반으로 광역상수도 강관의 노후도를 예측하는 모델을 개발하는 것이다. 데이터 학습을 위해 간접평가 데이터를 독립변수로, 직접평가 결과를 종속변수로 하는 노후도 예측모델을 개발하고자 한다. 이를 통해 그간 수행되었던 노후도 평가 방법에 있어 간접평가와 직접평가 방법의 한계점을 극복하고, 정확도 높은 노후도 예측을 통해 노후관 개량사업의 추진 효과성 제고와 조사과정의 안전, 단수문제를 해결하고자 한다.

이에, 제2장은 본 연구에서 해결하고자 하는 기존 간접평가와 직접평가 방식과 한계점을 선행연구와 실무경험을 바탕으로 제시한다.

제3장은 머신러닝 데이터 학습을 위해 분석 데이터의 특성을 확인하고, 데이터 전처리와 선행연구 고찰을 통해 데이터를 최적화한다.

제4장은 데이터 분석방법을 결정하고, 본 연구에 알맞은 머신러닝 알고리즘을 채택하여 정확도 높은 노후도 예측 모델을 구현하고 모델의 성능을 검증한다. 구현된 모델을 실제 노후관 개량사업에 적용하여 모델의 효과를 검증하였다.

제5장은 연구성과와 한계점을 논하고 향후 연구방향을 제시한다.

제2장 기존 노후도 평가 방법과 한계점

현재 상수도 관로의 노후도 평가는 상수관로 정밀조사 매뉴얼(환경부, 2020)에 따라 관 상태평가를 위한 정밀조사를 실시하고 있다. 본 장에서는 본 연구를 통해 해결하고자 하는 기존 노후도 평가방법을 살펴보고, 선행연구를 고찰하여 연구 방향을 설정하는 한편, 현행 방식의 한계점을 실무경험을 바탕으로 검토하여, 이를 개선하기 위한 연구 방법론을 탐색한다.

2.1 노후도 평가 관련 선행연구 고찰

김기태(2022)는 AI기반 상수도시설 개량 의사결정 모델 분석 연구에서 노후관 상태평가 데이터 결과를 활용하여 반복적인 학습으로 개량 의사결정 등급을 예측할 수 있는 인공신경망 알고리즘을 개발하였다. 2020년 한강유역의 노후관로 정비 기본계획에 의한 간접평가 데이터 12개 항목을 기반으로 데이터를 전처리하여 ANN 기반의 알고리즘을 적용함으로서 직접평가 결과값과 약 90%이상 일치되는 모델을 구현하였다. 동 연구는 관경 200mm 내외의 지방상수도 상태평가 데이터를 바탕으로 지역별(도별), 관종별 데이터 학습모델을 통해 평가하였다. 지방상수도는 대부분 블록(격자형) 단위로 운영되어 수지상식인 광역상수도와 관망 운영의 특성이 다르고, 소형관은 주철관, 비금속관이 주류를 이루는 반면, 중대형관인 광역상수도는 강관이 주류를 이루는 점을 고려하여 선행연구를 참고하여 금회 광역상수도 강관에 대한 노후도 예측모델 연구를 진행하였다. [23]

이윤호(2020)는 랜덤포레스트를 이용한 상수도관 부식연구에서 기계학습기법을 이용하여 관의 부식깊이 비율 예측모형을 개발하여 제시하였다. 수도권 3단계 광역상수도 283개 관로 구간 중 90%를 학습데이터로, 나머지 10%를 테스트 데이터로 활용하여 관의 부식깊이 비율을 예측하였다. 금회 연구는 동 연구를 참조하여 전국 데이터를 대상으로 부식깊이 비율 예측을 진행하였으나, RMSE가 높고 변수의 설명력이 낮아 노후도 평가 등급 분류 분석방법을 적용하였다. [22]

김주환(2019)은 노후 상수도관 관리실태와 진단을 통한 개선 연구에서 상수도관의 단계별 노후화 진행도를 제시하고, 노후화 단계별 관종별 영향인자를 제시하였다. 상수도관의 재질에 따라 노후화 메카니즘이 다른 특성을 나타내므로 기존의 영향인자들을 분석하여 관종별로 영향인자를 선정하였다. 관종이 다른 것은 관의 재질, 피복재의 종류, 관경이 서로 다르기 때문에 부식영향인자도 관종에 따라 서로 다르게 적용해야 함을 의미한다. [24]

김대웅(2021)은 인천시 상수도 관망의 노후도 평가에 있어 주성분 분석과 ANN 기법을 이용하여 상수도 관로의 노후도 점수평가를 수행하였다. 그 결과, 상수도기술진단에서 제시된 노후도 평가 10개의 항목을 간접평가 인자로 활용이 가능하다고 판단하였다. 따라서, 본 금회 연구에서 간접평가 데이터를 기반으로 독립변수를 선정시 활용하고자 한다. [25]

이슬민(2018)은 ANN-Clustering 기법을 이용한 상수관로 노후도 평가 및 분류 연구에서 기존의 점수평가법과 비교하여 9가지 평가항목을 모두 포함한 9-Clustering 모형과 평가항목 중 비교적 정량적 자료로 분류할 수 있는 4가지 항목만을 포함시킨 4-Clustering 모형

으로 구분하여 노후도 평가를 진행하였다. 연구 결과 입력변수를 조정한 4-Clustering 모형이 관로 노후도를 비교적 정확히 평가 및 분류하는 것으로 판단하였다. 이와 같이 모든 변수를 적용하는 것에 비해 적합도가 높은 변수를 적용할 때 모델의 성능이 향상되는 점을 고려하여 금회 연구에서도 독립변수를 최적화하고자 한다. [26]

이와 같이, 최근 상수도 관망의 노후화로 인한 다양한 사회적 문제가 발생하고 있고, 이를 개선하기 위한 다양한 연구가 진행되고 있다. 특히, ANN, 랜덤포레스트 등 머신러닝 기반의 연구를 통해 비용과 시간을 절감하고, 실제 직접조사에 준하는 결과값을 예측하기 위한 연구가 활발하다.

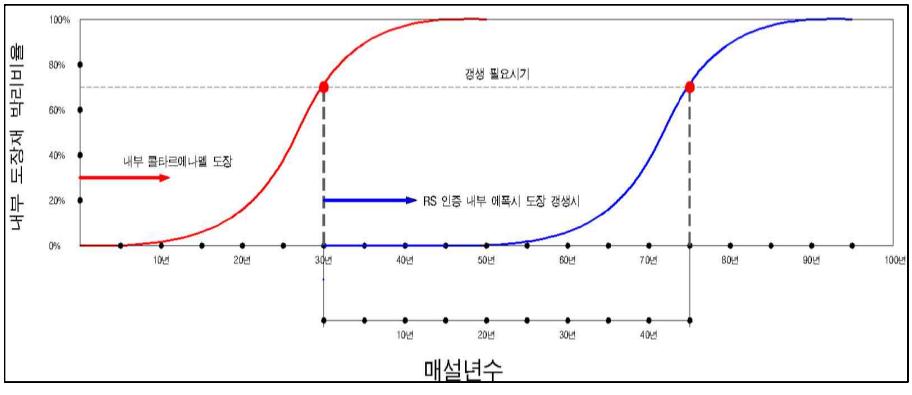
본 연구는 선행연구를 바탕으로 전국 광역상수도 데이터를 활용하여 머신러닝 기반으로 직접평가없이 간접평가 데이터를 활용한 정확도 높은 노후도 등급 분류 모델을 구현하고자 한다.

2.2 상수도 강관의 특성과 부식 매커니즘

상수도용 강관은 일반적으로 KS D 3565(상수도용 도복장 강관)에 따라 STWW 400 재질(열연강판을 가공하여 만든 원관)이며, KS에서 정한 화학성분과 기계적 성질을 만족하는 원관을 제작한 후 내면도장과 외면도장을 하는 도복장(코팅) 후 완성된 관을 사용한다.[8]

[그림 3] 상수도용 강관의 구조(그림출처 : 웰텍 주식회사 [9])

내면도장은 에폭시나 세라믹 계열의 수지로 도장(코팅)을 실시하는데 철재류로 된 원관이 내부의 물이나 산소, 이물질 등에 직접적으로 접촉되지 않도록 하기 위한 목적이다. 관로 매설 후 사용 중 복합적인 원인에 의해 내부 도장면의 국부적인 손상이 발생되면 내부를 흐르는 물의 수압과 전단력에 의해 도장면 전체로 손상부가 확대되어 원관이 직접 노출될 수 있기 때문에 노후도 평가 시 도장재의 박리비율 10%를 기준으로 개량대상 여부를 판단한다. 기존 광역상수도 대구경 강관에 대한 정밀조사(K-water연구원)에 따르면 도장재의 10%가 박리되면 급격하게 노후도가 진행되는 것으로 제시하고 있다. [10]



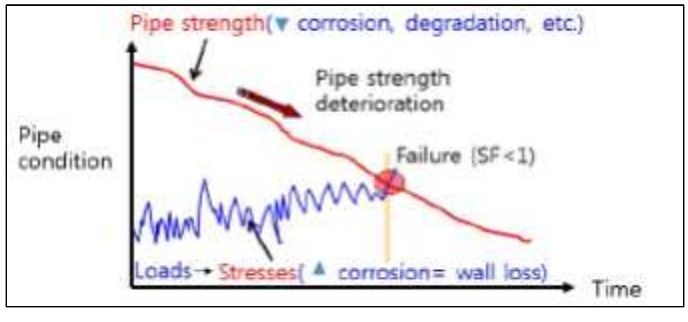
[그림 4] 매설년수별 내부 도장재의 박리비율

외면도장은 폴리에틸렌 계열 분말을 강관 외부에 융착하여 약 2mm 가량의 외부 피복을 형성하는데 이는 토양 부식성이나 함수율, 흙이나 돌에 의한 원관 손상, 전기적 요인에 의한 원관의 외부부식을 방지하기 위한 목적이다.

상수도용 강관의 원관은 철재류로서 산소, 물과 접촉하면 산화되어 부식되는 재질이며, 사용년수의 경과에 따라 다양한 요인에 의해 내·

외부에서 동시적인 부식이 발생할 수 있다. 이를 방지하기 위하여 상수도용 강관은 내부와 외부에 코팅을 하여 부식을 유발하는 인자와 강관의 관체간에 직접 접촉되지 않도록 하고 있다.

상수도 강관이 지중에 매설되면 매설환경과 운영조건에 따라 설치가 완료된 직후부터 점진적인 악화(Deterioration)가 진행된다. 내외부의 부식 진행정도에 따라 안전계수가 점차 감소하게 되며, 이론적으로 안전계수가 '1'이하가 되면 파손이 발생한다. 여기서, 안전계수가 '1'이라는 것은 구조물의 강도가 구조물의 사용상 요구되는 강도와 같음을 의미하는 것으로 구조적으로 안전한 상태를 유지하기 위한 최소한의 상태를 의미하므로 '최소 안전계수'라 표현하기도 한다. [4]



[그림 5] 상수도 강관의 부식 매커니즘

상수도 관은 대부분 지중에 매설하기 때문에 외부에는 되메우기를 한 토양과 접촉하게 되면서 토양의 부식성에 영향을 받는다. 정홍도(2012)는 누수가 발생한 관로 주변의 토양시료를 채취하여 화학 분석과 토양비저항을 측정하여 부식성을 평가하였다. 토양분석 결과 누수가 발생한 지점은 일반적인 토양에 비해 부식성이 상당히 높게 나타났으며, 토양의 부식성이 배관 전위 상승에 의하여 누수가 급격하게 진행된 것으로 판단하였다. 여기서 외면 부식 영향인자는 pH, 수분함량, 점토함량, 염소이온농도, 환산염이온농도, 토양비저항 값을 사용하였다.[11]

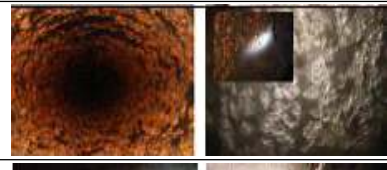
또한, 외면부식은 매설위치에 따라 차량하중과 충격하중의 영향을 받고, 매설깊이에 따라 토압(자중), 온도변화에 따른 신축, 동결하중의 영향을 받는다. 따라서, 본 연구에서는 상수도 강관의 외면 부식(노후화)에 영향을 미치는 인자를 고려하여 독립변수를 결정하고자 한다.

내면부식은 내부를 흐르는 유체의 수질특성과 수압에 의한 내부하중, 원수에 포함된 흙이나 이물질에 영향을 받는다. 한금석(2018)은 수돗물의 수질이 다른 2지역을 선정하여 수돗물에 대한 부식성지수와 수질변화에 따른 실험을 실시하였다. 그 결과, 수질부식성 지수가 관로의 내면부식에 영향이 있다는 것을 확인하였다.[12]

이를 종합한 강관 악화 메커니즘을 살펴보면 다음과 같다.

관 내면은 도장재 균열이나 국부적인 손상에 의해 부분적인 박리와 탈락이 발생하고, 도장재가 부분적으로 박리된 지점을 중심으로 관체의 전체 손상이 시작된다. 이후 도장재의 박리부가 점차 확대되고 수질특성과 수압, 유속에 의해 관 모체 손상이 전반적으로 진행되면서 관 두께가 점차 얇아져서 파손에 이르게 된다. [4]

관 외면은 시공단계의 결함, 충격이나 외부하중에 의한 외부 도장재의 손상이 발생되고, 토양부식성이나 물, 산소와 직접 접촉한 관의 모체는 시간이 경과하면서 점차 부식이 진행된다. 철의 부식 특성상 초기에는 국부적인 손상(Pin-Hole)으로 나타나지만, 부식 부위는 시간이 지날수록 점차 확대되어 관 모체 전반의 부식으로 연결되어 결국 파손에 이르게 된다. [4]

구분	단계		악화 단계 사진 (예시)	비고
내면	1단계	도장재 균열 Void 형성		도장재 악화 (점부식)
		용접부 국부부식		
	2단계	도장재 부분 박리 + 부식진행		관 모체 전체적 손상시작
	3단계	전면부식 진행 (pin-hole 발생)		구조적 악화 (응력 증가)
	4단계	파열		구조적 파손 (안전율 1이하)
외면	1단계	도장재 손상 (자갈, 시공)		국부적 손상
	2단계	국부적 부식 발생		다발적 손상
	3단계	파열		구조적 악화 (안전율 1이하)

[그림 6] 상수도 강관의 파손 진행단계 [5]

2.3 상수도 강관의 노후도 평가 방법

상수도 강관의 노후도 평가를 위한 정밀조사는 7단계로 구성된다.

먼저, 관로 정보수집 및 관망도 작성단계에서는 관로 상태를 파악하기 위한 관로시설 제원(관경, 매설년도, 매설깊이), 매설환경, 운영관리 이력 등에 대한 자료를 수집하고, 전산화 작업을 통해 관망도를 작성한 뒤 제원을 입력한다. 강관은 관경, 매설연수, 수질부식성, 토양부식성, 전기방식 유무, 매설깊이, 매설 위치 도로의 형태, 최대수압, 관로 사고 발생 건수, 분기관 설치 밀도 데이터를 수집한다.

이후 관로를 분할한다. 관로 분할은 관종, 매설연수, 매설환경 등의 차이에 따라 서로 다른 특성을 갖는 관로에 대하여 하나의 개별관로로 정의하고 최소단위로 평가한다.

다음으로 조사구간 설정을 위한 평가그룹을 설정한다. 평가그룹은 관종, 관경, 매설연수가 유사한 개별관로를 1.0km 단위로 그룹을 설정하여 조사구간을 설정하게 된다. 조사구간 설정이 완료되면 평가그룹별로 간접평가와 직접평가를 실시한다.

간접평가는 관로제원, 매설환경, 운영환경, 사고이력 등을 고려하여 관로의 상태를 간접적으로 측정하는 방법으로서, Ⅰ등급에서 Ⅲ등급까지 3가지 등급으로 노후도 평가등급을 측정한다.

직접평가는 간접평가 결과, 상태가 악화된 것(간접평가 3등급)으로 예측한 구간을 현장 직접조사 구간으로 선정하여 조사를 시행한다. 관체를 굴착하여 외면의 상태조사를 실시하고, 관 절단 후 관 내부에 진입하거나 부단수 내시경 등의 방법으로 관 내부의 상태조사를 실시한다. 조사결과를 분석하여 Ⅰ등급(양호), Ⅱ등급(세척), Ⅲ등급(갱생), Ⅳ등급(구조보강), Ⅴ등급(교체)의 노후도 등급이 결정된다. [5]

직접평가까지 완료되면 관체의 건전성 평가를 거쳐 관의 악화원인과 상태에 따라 최종 노후관 개량방법을 결정하게 된다. 노후관 개량방법에 따라 투자사업비, 우선순위 결정, 시공방법이 결정된다.

[표 2] 상수도 관로의 노후도 평가 단계

단계	구 분	내 용
1단계	관로정보 수집 관망도 작성	- 관로 상태를 파악하기 위한 관로시설 제원, 매설환경, 운영관리 이력 등에 대한 자료를 수집 - 전산화 작업을 통해 관망도를 작성하고 제원입력
2단계	관로분할 (개별관로)	- 관로제원, 관리조건, 운영조건 등에 따라 평가를 위한 최소관로 단위(개별관로) 설정
3단계	조사구간 설정 (평가그룹)	- 관종, 관경, 매설연수가 유사한 개별관로를 1.0km 단위로 그룹을 설정하여 조사구간 설정
4단계	간접평가	- 관로제원, 매설환경, 운영환경, 사고이력 등을 고려하여 관로의 상태를 평가
5단계	직접평가	- 간접평가 결과 상태가 악화된 것으로 예측된 구간을 현장 직접조사 구간으로 선정 - 관체 내외면 상태를 직접조사하고 정량적으로 평가
6단계	관체 건전성 평가	- 직접조사 대상지점의 관체의 물리적 특성(강도, 화학조정, 금속조직 등)을 조사하여 건전성을 평가
7단계	개량방안 수립	- 관의 악화원인과 상태에 따라 개량방법 결정 * 소구경은 부식깊이, 중대구경은 안전계수로 결정

2.4 상수도 강관의 간접평가 방법

상수도 강관의 간접평가는 점수평가 방법을 적용한다. 점수평가 방법은 관 파손 영향인자들에 대하여 가중치를 적용한 점수를 부여하여 관의 상태를 예측하는 방법으로 평가 결과가 점수로 결정된다. 강관의 간접평가 항목은 총 10가지로 구분된다. [5]

[표 3] 상수도 강관의 간접평가 항목

구 분	구 분	내 용	가중치
관경	공칭관경 (㎜)	내면 용접유무 관 두께	0.098
매설 연수	매설후 경과년수	경년화	0.255
수질 부식성	L.I. (Langelier Index)	관내면 부식	0.041
토양 부식성	토양비저항 (Ω-㎝)	관외면 부식	0.059
전기 방식	방식전위 (-850 ~ -2,500㎷)	관외면 부식 (3년간 측정 자료)	0.118
매설 깊이	지상에서 관상단까지의 매설깊이	외부 사하중(토압)	0.028
도로 형태	차선수 및 용도에 따라 구분	외부 활하중(차량하중)	0.037
최대 수압	평가구간의 최대수압 (㎏f/㎠)	내부 하중(최대 수압)	0.083
파손 건수	최근3년간 파손건수 (건/㎞/yr)	자연적으로 발생한 파손 (누수)	0.194
밸브/ 분기수	관로상의 접속관/분기관 수 (개수/㎞)	부속시설 연결 천공 등에 의한 관체 손상	0.087

관 상태 간접평가 방법은 관종별 평가항목 조건값에 가중치를 곱한 값을 가중치 합으로 나누고 매설년수에 따른 관 상태 보정계수를 곱해 간접평가 점수를 산정한다. 간접평가 점수가 60점 이상인 경우 Ⅰ등급(상태양호), 40점~60점 미만인 경우 Ⅱ등급(노후진행), 40점 미만인 경우 Ⅲ등급(노후추정)으로 판단한다. [5]

[표 4] 상수도 강관의 간접평가 점수 산정방법

$$\text{관 상태 간접평가 점수} = \left(\frac{\sum f_{p_i c_i} \cdot w_{p_i c_i}}{\sum w_{p_i c_i}} \right) \times DP_{\max_i}$$

여기서, $f_{p_i c_i}$ = 관종별 평가항목의 조건값

$w_{p_i c_i}$ = 관종별 평가항목의 가중치

$DP_{\max_i}$ = 매설연수별 관상태 보정계수

매설연수	관 상태 보정계수
0 ~ 10년 이하	100
10 ~ 20년 이하	90
20 ~ 30년 이하	80
30년 초과	75

[표 5] 간접평가와 직접평가 등급 비교표

간접평가점수	구 분	내 용	
		등급	개량방안
60점 이상	Ⅰ(노후진행 소)	Ⅰ	상태양호 또는 세척
		Ⅱ	
40 ~ 60점	Ⅱ(노후진행 중)	Ⅱ	세척 또는 갱생
		Ⅲ	
40점 미만	Ⅲ(노후진행 대)	Ⅲ	갱생, 교체 (방식, 보강, 교체)
		Ⅳ, Ⅴ	

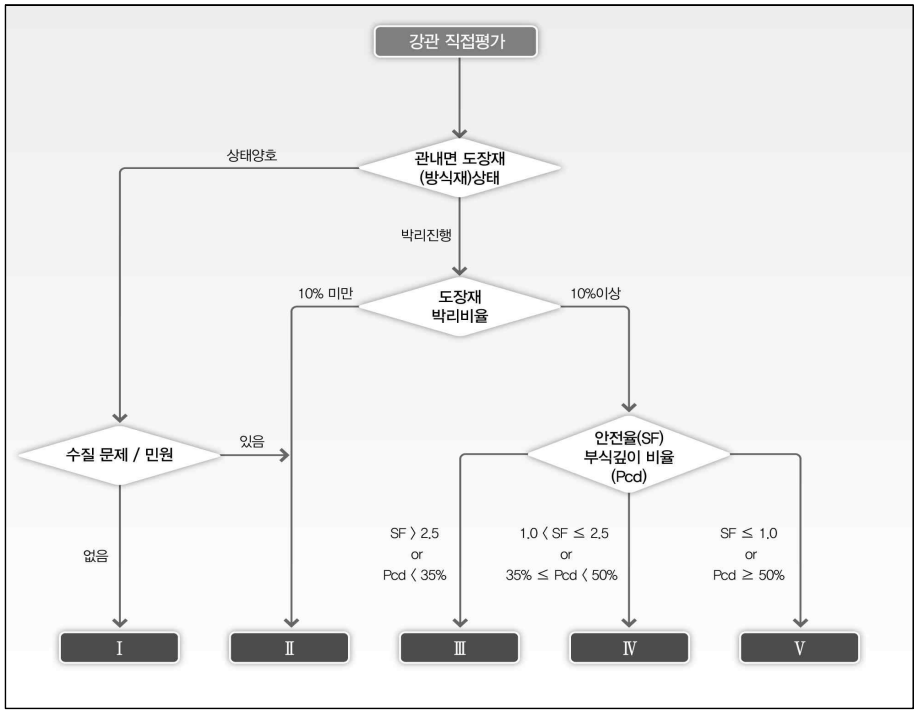
2.5 상수도 강관의 직접평가 방법

상수도 강관의 직접평가는 표본굴착 또는 비파괴 조사장비를 이용하여 관의 상태를 직접확인하여 평가하는 방법이다. 간접평가 결과 Ⅲ등급(필요시 Ⅱ등급)으로 평가된 조사구간을 직접평가 조사지점으로 선정한다. 조사지점은 개별관로 그룹 중 점수가 가장 낮은 구간을 대상으로 한다. 평가항목은 표6과 같다.[5]

[표 6] 상수도 강관의 직접평가 항목

구 분		측정방법	비 고
관 제원	관경	줄자	굴착 후 조사
	관두께	버니어캘리퍼스	
내부 슬라임	분포면적	육안조사	단수 후 조사
	축적두께	버니어캘리퍼스	
내면 도장상태	도장재 상태	육안조사	
	도장재 손상비율	육안조사	
외면 도장상태	도장재 상태	육안조사	굴착 후 조사
	도장재 손상비율	육안조사	
내부 결절상태	분포비율	육안조사	단수 후 조사
	결절크기	버니어캘리퍼스	
내면 부식상태	침식면적	육안조사	
	최대부식깊이	깊이 게이지	
외면 부식상태	침식면적	육안조사	굴착 후 조사
	최대부식깊이	깊이게이지	
관재료 건전성	물리적 강도	인장강도 시험	관체 절단 후 조사
	화학적 조성	금속 현미경	

직접평가 항목에 따라 조사가 완료되면 그림 7의 흐름도에 따라 노후도 평가등급을 결정한다. 우선, 관 내면 도장재 상태를 판단하여 도장재의 상태가 양호한 경우 Ⅰ등급으로 분류하고, 관 내면 도장재의 박리가 진행된 경우 도장재의 박리비율 10%를 기준으로 10%미만인 경우 Ⅱ등급으로 분류한다. 도장재의 박리비율이 10%이상인 경우에는 부식 진행정도를 측정한 결과에 따라 부식깊이 비율이 최초 공칭 관두께 대비 얼마나 감소하였는지를 1차적으로 판단하여 부식깊이 비율 35%, 50%를 기준으로 직접평가 등급을 결정한다. 또는 관에 작용하는 응력대비 잔존 인장강도의 비율로 안전계수를 계산하여 1.0, 2.5를 기준으로 직접평가 등급을 결정한다. 위 두가지 방법을 동시에 평가를 하여 가장 악조건의 결과를 최종 직접평가 등급으로 결정한다.



[그림 7] 강관 직접평가 등급 판정 흐름도 [5], [13]

직접평가 등급별 상태조건에 따라 개량방안을 결정한다.

Ⅰ등급은 상태가 양호하여 계속 사용한다. Ⅱ등급은 내부에 슬라임이나 침전물이 축적되어 수질민원이 발생하는 경우로서 관 세척을 실시한다. Ⅲ등급은 내부 도장재의 박리비율이 10%이상 진행되고, 부분적인 부식이 진행되는 단계로서 비구조적 갱생을 통해 개량사업을 진행한다. 비구조적 갱생이란 관 내부의 이물질이나 녹을 스크래핑을 통해 제거하고, 새로운 내부 도장재를 도포하는 갱생방법이다. Ⅳ등급은 부식깊이 비율이 35%~50%까지 상당부분 진행된 경우로서 구조적 갱생을 진행한다. 구조적 갱생은 부식깊이가 깊은 지점에 구조재로 보강하거나 관 내부에 기존 관보다 약간 작은 관 또는 튜브를 삽입하는 형태의 갱생방법이다. Ⅴ등급은 관을 완전히 교체하는 방법이다.

[표 7] 직접평가 등급에 따른 관의 상태와 개량방법

직접평가 등급	상태평가	개량 방법
I	• 내면 도장재 상태양호 및 표면에 슬라임 축적 ➜ 수질영향 없음 또는 수질민원 없음	상태양호
II	• 내면 도장재가 부분적으로 박리되고 표면에 슬라임 축적 ➜ 내면 도장재 박리비율 10%이하 또는 수질민원 발생 (부식깊이 비율 35% 미만)	세척
III	• 내면 도장재 방식기능 저하 ➜ 내면 도장재 박리비율 10%이상 내압 또는 외압에 대한 안전계수 : SF > 2.5 (부식깊이 비율 35% 미만)	비구조적 갱생
IV	• 내면 도장재 방식기능 저하 ➜ 박리비율 10% 이상 내압 또는 외압에 대한 안전계수 : 1.0 < SF ≤ 2.5 (부식깊이 비율 35~50%)	구조적 갱생
V	• 내면 도장재 라이닝 방식기능 상실 ➜ 박리비율 10%이상 (결절성장) 내압 또는 외압에 대한 안전계수 : SF ≤ 1.0 (부식깊이 비율 50% 이상)	관 교체

2.6 기존 상수도 강관의 노후도 평가의 한계점

2.6.1 간접평가 방법의 한계점

2.6.1.1 특정지점 조사로 인해 입력자료에 따라 정확성이 다름

간접평가는 관로 분할된 구간 중 특정 지점을 조사하고 그 결과를 가중평균하여 해당 그룹의 간접평가 점수를 도출한다. 하지만, 길이가 긴 관로 중 조사가 가능한 임의의 지점을 선정하고 있기 때문에 조사지점을 어디로 결정하느냐에 따라 간접평가 결과가 달라지게 된다. 조사지점에 따라 달라지는 매설깊이와 도로형태는 관로 상부의 토압과 차량하중에 영향을 미치게 되고, 토압과 차량하중은 관로에 작용하는 응력에 결정적인 인자가 된다. 또한, 토양부식성은 매설위치에 따라 토양의 특성이 달라지게 되는데 이는 관로 외면부식에 영향을 미친다. 최대수압은 관로 매설지점의 높낮이에 따라 달라져 관로 내면부식에 영향을 미치게 된다. 위와 같이 조사지점에 따라 입력자료가 달라지게 되고, 입력자료에 따라 간접평가의 정확성이 달라진다.

따라서, 모든 구간을 세부적으로 조사하는 것이 최선의 방법이지만 조사비용과 시간이 지나치게 많이 소요되기 때문에 조사의 용이성과 운영여건을 고려하여 대표성을 갖는 지점을 선정하고 있는 실정이다.

2.6.1.2 간접평가 결과만으로 노후관 개량방안 결정 곤란

간접평가는 예측된 결과이므로 그 결과만 의존해서 개량방안을 마련하기는 어렵다. 따라서 대부분 간접평가 방법은 문제 관로를 골라내는 screen으로 활용하고, screen을 통해 걸러진 관로들을 대상으로 연차별 개량계획 또는 우선순위를 결정하는데 주로 활용한다. [5]

이에, 국가수도기본계획이나 연차별 투자 우선순위 결정, 개량사업에 소요되는 개략사업비의 산정은 간접평가 결과를 활용하고 있다. 직접평가는 대부분 기본 및 실시설계 단계에서 시행하므로 계획을 수립하는 단계에서는 직접평가 결과를 활용하는 것이 곤란하기 때문이다.

하지만, 간접평가 결과 등급에 따라 직접평가 등급을 예측하기가 어렵고, 실제 결과에서도 차이가 발생하고 있는 실정이다. 표8은 본 연구에서 분석데이터로 활용한 데이터의 간접평가 결과와 직접평가 결과를 비교한 결과이다.

[표 8] 분석 데이터의 간접, 직접평가 분포표

직접평가 / 간접평가	Ⅰ등급	Ⅱ등급	Ⅲ등급	Ⅳ등급	합 계
Ⅰ(노후진행 소)	264	210	-	-	474 (60.2%)
Ⅱ(노후진행 중)	-	20	163	64	247 (31.4%)
Ⅲ(노후진행 대)	-	3	55	8	66 (8.4%)
합 계	264 (33.5%)	233 (29.6%)	218 (27.7%)	72 (9.2%)	787 (100%)

간접평가 Ⅱ등급인 경우에 노후도가 높은 직접평가 Ⅳ등급으로 판정된 경우가 많기 때문에 간접평가 결과만으로 개량방안을 결정하기 어려울 뿐 아니라, 직접평가 대상구간 누락이 발생할 수 있는 한계점이 있다. 직접평가 등급에 따라 소요되는 사업비나 우선순위가 다르기 때문에 현재의 노후도 평가 체계에서는 사업비나 우선순위 선정이 제대로 수행되기 어려운 실정이다.

D1,000mm를 기준으로 각 개소를 100m로 간주한 후 실제 간접평가와 직접평가 등급에 따라 소요되는 사업비를 산출하면 44,853백만원이 소요된다. 하지만 간접평가 등급으로 직접평가 결과를 정확하게 확인할 수 없으므로 표5에 따라 간접평가 Ⅰ등급은 Ⅰ~Ⅱ등급으로, Ⅱ등급은 Ⅱ~Ⅲ등급으로, Ⅲ등급은 Ⅲ~Ⅴ등급으로 가정하고 소요사업비를 정량적으로 비교하였다. 그 결과, 간접평가 결과 상태가 모두 가장 양호할 것으로 예측될 경우(간접평가 Ⅲ등급인 경우 직접평가를 모두 Ⅲ등급으로 추정)에는 10,415백만원, 상태가 모두 불량할 것으로 예측한 경우에는 49,146천원으로 사업비 편차가 있는 것으로 확인되었다.

즉, 실제에 비해 적게는 23.2%로 과소 책정될 수 있고, 많게는 109.6%로 과다하게 책정되는 등 수십억에서 수백억의 총사업비 편차가 발생할 수 있다.

따라서, 사업계획 수립 단계에서 현행 간접평가보다 정확도가 높은 노후도 예측을 통해 사업비 산정과 우선순위 결정의 효과성을 제고하기 위한 개선이 필요하다.

2.6.1.3 간접평가와 직접평가의 평가등급 체계 상이

간접평가는 점수평가법에 따라 Ⅰ~Ⅲ등급까지 3가지 등급체계로 평가하고 있으며, 직접평가는 Ⅰ~Ⅴ등급까지 5가지 등급체계로 평가하고 있다. 즉, 간접평가와 직접평가는 평가등급 체계가 상이하다. 표5와 같이 간접평가 Ⅱ등급인 경우 조견표에 따라 직접평가는 Ⅱ~Ⅲ등급으로 간주하고 있으나, 실제 결과는 표8과 같이 Ⅱ~Ⅳ등급으로 확인되었다. 간접평가 Ⅲ등급인 경우 Ⅲ~Ⅴ등급으로 간주하고 있으나, 실제 결과는 Ⅱ~Ⅴ등급으로 다양하게 분포하는 것을 확인하였다.

여기서 예상되는 문제는 각 평가등급에 따라 소요되는 사업비에 차이가 크다는 점이다. 수도시설 운영비 및 개략공사비 산정기준(환경부)에 따르면 직접평가 Ⅱ등급에 적용하는 세척의 평균 단가는 150천원/m, Ⅲ등급에 적용하는 비구조적 갱생공법의 평균 단가는 D1,000mm를 기준으로 1,316천원/m, 직접평가 Ⅳ등급에 적용하는 구조적 갱생은 2,018천원/m, 직접평가 Ⅴ등급에 적용하는 노후관 교체는 2,624천원/m이다. 직접평가 등급에 따라 결정되는 개량방법별로 평균 단가의 차이가 크기 때문에 사업계획단계에서 정확도 있는 등급예측이 매우 중요하다.[14]

하지만, 현행 간접평가 결과를 바탕으로 투자우선순위와 사업비를 예측할 경우 직접평가 등급과 평가등급 체계가 상이함에 따라 정확도 있는 사업계획을 수립하는 것이 매우 곤란한 실정이다.

사업계획 단계에서 사업비를 잘못 산정하여 과소하게 산출할 경우, 향후 설계나 시공단계에서 사업비의 대규모 증액이 예상되고 이는 초기 투자비율 변동에 따른 사업성을 저하시킬 수 있고, 총사업비의 변동으로 추가적인 행정력 소요와 계약상대자와의 클레임으로 연결될 가능성이 매우 높다. 반대로 사업비를 과다하게 산출할 경우, 국가수도기본계획 등 국가계획의 신뢰성 문제로 연결될 가능성이 있다.

따라서, 현행 간접평가와 직접평가의 평가등급 체계가 상이한 부분에 대하여 본 연구를 통해 개선이 필요하다.

[표 9] 관경별, 노후관 등급별 단위 공사비

(단위: 원)

관경 (mm)	단위	Ⅱ등급 관세척	Ⅲ등급 비구조적 갱생	Ⅳ등급 구조적 갱생	Ⅴ등급 (교체 또는 신설)
D700	m	50,000 ~ 500,000	971,555	1,689,000	1,706,151
D800	m		1,193,646	1,712,000	1,876,254
D900	m		1,285,919	1,856,000	2,100,165
D1000	m		1,316,187	2,018,000	2,624,908
D1100	m		1,420,793	2,312,000	2,940,649
D1200	m		1,491,875	2,377,000	3,131,289
D1350	m		1,605,326	2,666,000	3,438,858
D1500	m		1,699,580	3,138,000	3,869,520
D1650	m		1,831,910	3,417,000	4,366,376
D1800	m		1,939,479	3,851,000	4,903,367
D2000	m		2,123,354	4,919,000	5,400,464
D2200	m		2,375,211	5,781,000	7,125,855
D2300	m		2,545,661	6,067,000	8,561,513

* 단 , 구조적 갱생은 관 자재비를 포함한 비용으로 반영함

2.6.2 직접평가 방법의 한계점

직접평가 방법은 간접평가 결과 상태등급 Ⅲ등급(필요시 Ⅱ등급)으로 평가된 조사구간을 직접평가 조사지점으로 선정한다. 선정된 구간 중 운영상황, 인허가 여건, 단수가능여부를 종합적으로 판단하여 최종 조사지점을 결정하고, 관 외면의 굴착 조사를 실시한다. 관 내면은 단수가 가능할 경우 관체를 절단한 후 조사자가 내부로 진입하여 내면 조사를 실시하고, 단수가 불가능할 경우 부단수 내시경을 통해 관 내면의 상태를 평가한다.

2.6.2.1 밀폐·협소공간 조사에 따른 위험도 상승

관 외면 조사는 도로 등의 관로 매설지점을 굴착하여 조사를 실시한다. 굴착지점은 매우 협소한 공간이며, 대부분 도로상 굴착으로 인해 차량이 통행하면서 2차 사고의 위험도가 상존한다.

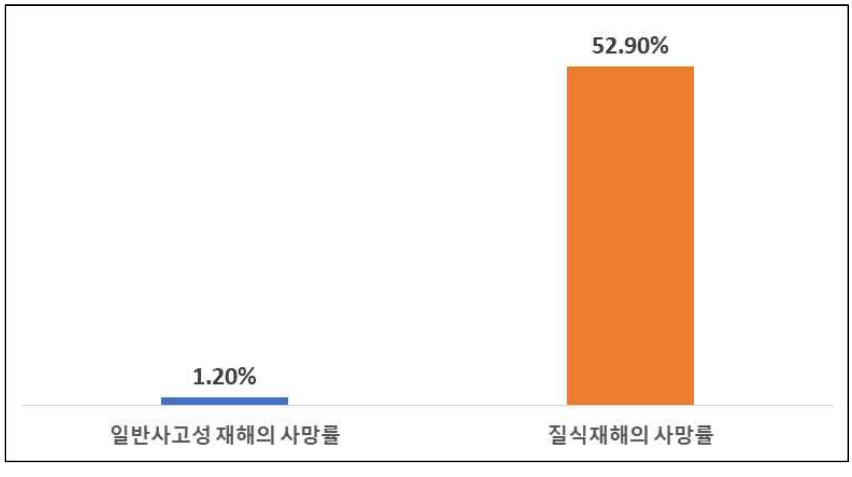
관 내면 조사는 직경 700mm 이상 관로에서 단수가 가능할 경우 가능한데, 조사자는 몸을 움츠리고 밀폐된 관 내부에 들어가 장시간 조사를 해야하기 때문에 움직임의 제한, 산소 결핍, 유해가스로 인한 질식 등 밀폐공간 내 안전사고의 위험에 노출된다.

관 내면조사 전경	관 외면조사 전경

[그림 8] 관 노후도조사 전경

밀폐공간 질식재해예방 안전작업 가이드(고용노동부, 안전보건공단)에 따르면 질식재해가 발생하면 2명 중 1명은 중대재해(사망)가 발생하는 것으로 조사하였다. 일반 사고성 재해에서 사망자가 차지하는 비율은 전체 재해자의 1.2%이나, 질식재해에서 사망자가 차지하는 비율은 52.9%에 이르고 있어 2명 중 1명이 사망할 만큼 치명적이다.[15]

특히, 전체 질식재해 사망자 중 44%는 건설업이며, 건설업 중 콘크리트 양생작업, 오폐수처리시설, 맨홀 공사현장, 배관공사에서 많이 발생하는 것으로 조사하였다. 노후관 직접조사는 건설업 중 맨홀 또는 배관공사에 해당하는 것으로 밀폐공간 질식 재해의 위험성이 매우 높다는 것을 확인할 수 있다.

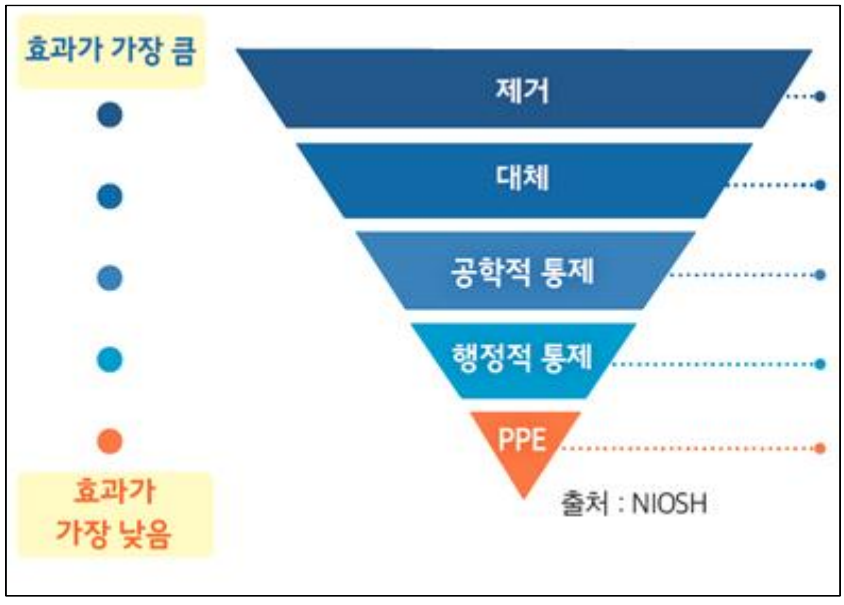


[그림 9] 일반사고성 재해와 질식재해의 사망률

이처럼 밀폐공간에서 행해지는 노후관 직접조사는 예기치 못한 위험성을 내포하고 있어 조사시 고도의 안전관리가 요구된다. 미국 국립 직업안전위생연구소(NIOSH)에 따르면 위험을 제거하는 가장 효과가 큰 방법은 위험요인 자체를 제거하는 것이다. 즉, 밀폐공간 조사를

하지 않더라도 그에 준하는 조사 결과를 얻는 것이 가장 효과가 큰 위험관리 방법이라 할 수 있다. [16]

따라서, 본 연구는 밀폐·협소공간의 조사없이 실제 조사에 준하는 결과를 얻을 수 있도록 개선방향을 도출하고자 한다.



[그림 10] 안전사고 예방을 위한 효과적 관리방안

2.6.2.2 조사시 단수로 인한 국민 불편과 조사지점 제한

직접조사를 수행하기 위해서는 관체를 절단하고 관 내부에 조사자가 진입하여 조사를 실시하여야 한다. 일반적으로 직경 D700mm 이상 관로까지 조사자가 내부로 진입할 수 있는데 직경별로 단수가 발생할 경우 D700mm 기준 144,928명에서 D2,300mm 기준 2,318,841명까지 대규모의 단수가 발생할 수 있다.

이처럼 직접조사를 위해서는 단수가 수반되기 때문에 단수없이 실시할 수 있는 부단수 내시경 조사를 실시하거나 인근의 단수가 가능

한 구간의 직접조사 결과를 준용하여 활용하고 있는 실정이다. 단수가 수반되기 때문에 제한적인 위치만 조사를 시행함에 따라 조사 결과의 대표성 부족 문제가 발생하게 된다.

실제 수도시설 안정화 전략계획 수립 용역(K-water)에 따르면 전국 광역 및 공업용수도의 노후도 평가를 위해 간접평가 약 14,000여 건을 수행하였으나, 실제 직접조사를 수행할 수 있는 구간은 전체의 2.3%인 326건으로 확인되었다. 운영 중 단수를 할 수 있는 구간은 복선화 관로가 설치되어 대체 공급이 가능하거나 다른 시설과 연계운영이 가능한 구간, 충분한 배수지 용량이 확보되어 한시적으로 단수를 하더라도 물 공급에 차질이 없는 구간으로 매우 제한적이다.[17]

따라서, 본 연구는 단수로 인한 국민불편과 조사지점의 제한없이 직접조사에 준하는 노후도 평가 결과를 얻고자 한다.

[표 10] 관경별 단수시 예상 단수인구 현황

관경 (mm)	경제적 유량 (㎥/일)	예상 단수인구 (명)
D700	50,000	144,928
D800	60,000	173,913
D900	80,000	231,884
D1000	90,000	260,870
D1100	120,000	347,826
D1200	140,000	405,797
D1350	220,000	637,681
D1500	260,000	753,623
D1650	300,000	869,565
D1800	360,000	1,043,478
D2000	500,000	1,449,275
D2200	600,000	1,739,130
D2300	800,000	2,318,841

2.7 기존 노후도 평가 한계점 개선을 위한 연구방향

기존 노후도 평가 방법에 있어 간접평가와 직접평가의 한계점을 살펴보았다. 간접평가는 ① 특정지점 조사로 인해 입력자료에 따라 정확성이 달리지는 점 ② 간접평가 결과만으로 개량방안을 결정하는 것이 곤란한 점 ③ 간접평가와 직접평가의 평가등급 체계가 상이한 점에서 한계가 있다. 직접평가는 ① 밀폐·협소공간 조사에 따른 조사시 위험도가 매우 높은 점 ② 조사를 위한 단수 수반으로 국민 불편이 발생하며, 조사지점이 제한되는 점에서 한계가 있다.

이에 본 연구는 직접평가를 실시하지 않더라도 직접평가 결과에 근접한 정확도 높은 노후도 평가 등급을 예측하고자 한다.

이를 위해 데이터 분석 기반의 머신러닝 기법을 활용하여 직접평가 등급을 예측할 수 있도록 예측 모델을 개발하고, 이를 실무에 적용하여 예산절감과 향후 후속사업의 활용성을 제고하고자 한다.

[그림 11] 연구방향 (AS-IS - TOBE)

제3장 데이터 특성 분석을 통한 변수 최적화

본 장에서는 분석에 사용한 데이터의 특성을 살펴보고, 데이터의 특성과 중요도를 고려하여 노후도 예측 모델에 적용할 독립변수와 종속변수를 채택하고자 한다.

3.1 노후도 평가 데이터 수집

데이터 기반의 머신러닝 기법 적용을 위해 간접평가와 직접평가 데이터를 수집하였다. 데이터는 2011년 이후 시행된 11개의 광역 및 공업용수도 노후관 개량사업 기본 및 실시설계 단계에서 수행된 간접평가 결과와 직접평가 결과 787개×12개 항목의 데이터를 수집하였다.

	DIA	Year	LI	Soil	Electronic	Depth	Load	Pressure	Accident	Branch	Indirect	Corruption
0	1100	36	-2.09	17907	1.00	1.77	1.0	4.22	0.0	1.0	48.73	Ⅳ
1	1100	36	-2.09	12252	1.00	1.85	1.0	4.22	0.0	0.0	48.73	Ⅳ
2	1100	36	-2.09	12252	0.35	2.17	1.0	4.22	0.0	1.0	48.20	Ⅳ
3	1100	36	-2.09	12252	1.00	2.78	1.0	4.22	0.0	7.0	41.68	Ⅳ
4	1100	36	-2.09	12252	1.00	2.71	1.0	4.22	0.0	1.0	48.20	Ⅳ

[그림 12] 분석데이터 현황

광역 및 공업용수도 시설은 전국 48개 시설을 운영중이며, 향후 연구 모델의 범용적인 적용을 위해서 그간 전국에서 수행한 데이터를 수집하여 모델의 학습데이터로 활용하였다.

수집한 데이터는 간접평가 항목 10가지(관경(DIA), 매설연수(Year), 수질부식성(LI), 토양부식성(Soil), 전기방식 유무(Electronic), 매설깊이(Depth), 매설위치(Load), 내부압력(Pressure), 사고이력(Accident), 분기점 밀도(Branch))와 간접평가 점수(IN-Direct) 등 11가지 항목을

수집하였으며, 직접평가 항목은 노후관 개량대상여부(Ⅰ~Ⅱ등급인 경우 계속사용, Ⅲ~Ⅴ등급인 경우 개량대상으로 분류), 직접평가 등급에 대한 2가지 데이터를 수집하였다. 각 데이터의 기초 통계량은 다음 표와 같다.

```
     DIA           Year            LI              Soil          Electronic         Depth            Load
 Min.   : 600   Min.   : 1.00   Min.   :-2.450   Min.   : 2373   Min.   :0.0000   Min.   : 0.700   Min.   :0.0000
 1st Qu.:1200   1st Qu.:11.00   1st Qu.:-1.420   1st Qu.: 8192   1st Qu.:0.7000   1st Qu.: 2.100   1st Qu.:0.7000
 Median :1650   Median :15.00   Median :-0.980   Median :12010   Median :1.0000   Median : 2.590   Median :1.0000
 Mean   :1606   Mean   :21.38   Mean   :-1.154   Mean   :12547   Mean   :0.7985   Mean   : 2.831   Mean   :0.8644
 3rd Qu.:2000   3rd Qu.:35.00   3rd Qu.:-0.850   3rd Qu.:15000   3rd Qu.:1.0000   3rd Qu.: 3.225   3rd Qu.:1.0000
 Max.   :2600   Max.   :51.00   Max.   :-0.070   Max.   :90478   Max.   :1.0000   Max.   :31.090   Max.   :1.0000
    Pressure         Accident            Branch          Corruption
 Min.   : 0.300   Min.   :0.000000   Min.   :  0.000   Length:787
 1st Qu.: 4.270   1st Qu.:0.000000   1st Qu.:  0.800   Class :character
 Median : 5.070   Median :0.000000   Median :  2.160   Mode  :character
 Mean   : 5.784   Mean   :0.001128   Mean   :  3.481
 3rd Qu.: 7.100   3rd Qu.:0.000000   3rd Qu.:  2.950
 Max.   :16.800   Max.   :0.338800   Max.   :333.330
```

```
'data.frame':   787 obs. of  11 variables:
 $ DIA       : int  1100 1100 1100 1100 1100 900 900 900 900 900 ...
 $ Year      : int  36 36 36 36 36 36 36 36 36 36 ...
 $ LI        : num  -2.09 -2.09 -2.09 -2.09 -2.09 -2.09 -2.09 -2.09 -2.09 -2.09 ...
 $ Soil      : int  17907 12252 12252 12252 12252 12252 6597 6597 6597 6597 ...
 $ Electronic: num  1 1 0.35 1 1 1 1 1 1 1 ...
 $ Depth     : num  1.77 1.85 2.17 2.78 2.71 2.34 2.27 1.66 1.54 1.62 ...
 $ Load      : num  1 1 1 1 1 1 1 1 1 0.7 ...
 $ Pressure  : num  4.22 4.22 4.22 4.22 4.22 4.22 4.22 4.22 4.22 4.22 ...
 $ Accident  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ Branch    : num  1 0 1 7 1 2 2 6.89 1 1 ...
 $ Corruption: chr  "IV" "IV" "IV" "IV" ...
```

[그림 13] 분석 데이터의 기초 통계량

다만, 데이터 수집은 종속변수로 사용되는 직접평가가 수행된 경우에 한해 수집이 가능하여 금회 연구에서는 787개 데이터를 확보하였으나, 분석의 신뢰도 향상을 위해서는 계속적인 데이터 축적이 필요하다.

3.2 노후도 평가 데이터의 특성 분석

수집된 데이터의 특성을 살펴보았다. 독립변수로 사용할 간접평가 항목 10가지 데이터의 특성과 필요성, 중요도를 검토하여 연구모델의 성능 향상에 기여하는 독립변수를 채택한다.

3.2.1 관경(Diameter)

관경은 노후도 평가에서 매우 중요한 요소이다. 광역 및 공업용수도 건설사업에서 관경은 용수수요량과 내외부 압력, 소요사업비를 고려하여 최적 관경으로 결정하는데, 결정된 관경에 따라 관의 두께가 달라진다. 관경이 클수록 관의 두께도 두껍기 때문에 동일한 부식환경에서 같은 수준의 부식이 발생하였다 하더라도 관경이 크고, 관의 두께가 두꺼운 관로에서 안전율이 더 높다. 또한, 표 11과 같이 노후도 직접평가에서 내압에 의한 원주방향응력 계산시 관의 직경은 중요한 인자로 적용되고 있다.

[표 11] 내부압력에 의한 원주방향응력 산정 공식

$$\sigma_h = \frac{P_{total\,\mathrm{total}} \times d}{2\mathrm{t}}$$

여기서, σ_h = 내부압력에 의한 원주방향응력 (kgf/cm²)

t = 관 두께 (cm)

d = 관 내경 (cm)

P_{total} = 총 내부압력 (kgf/cm²) (P_w : 운전수압, P_s : 수충격압)

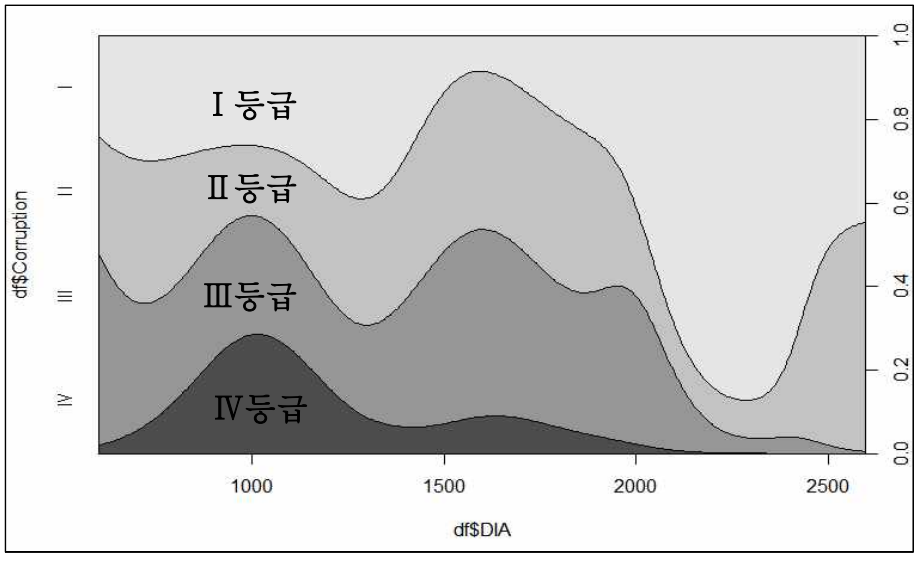
그림 14와 같이 관경이 작은 관일수록 직접평가 등급에서 Ⅲ등급 또는 Ⅳ등급으로 판정되는 비율이 높았고, 관경이 클수록 Ⅰ~Ⅱ등급의 비중이 높았다. 이러한 결과는 수집된 데이터에 따라 달라질 수 있기 때문에 분석에 활용된 데이터 중 30년 이상 경과된 관로를 대상

으로 평균 직경과 평균 매설년수를 비교하였다. 평균 매설년수는 비슷한 수준으로 확인되었으나, Ⅲ등급으로 평가된 관로의 평균 관경은 1,499mm, Ⅳ등급으로 평가된 관로의 평균 관경은 1,359mm로서 관경이 클수록 직접평가 등급이 상대적으로 양호한 것을 확인하였다.

이와 같이 노후도 평가 등급은 관경의 영향이 있는 점, 직접평가 등급 산정시 관경이 중요한 인자인 점, 관경 데이터가 600mm~2,600mm까지 다양하게 분포되어 있는 점을 고려하여 본 연구에서는 관경 데이터를 독립변수로 채택하여 활용하고자 한다.

[표 12] 직접평가 등급별 평균 직경 현황

직접평가 등급	데이터 수	평균직경	평균 매설년수
Ⅲ등급	197	1,499mm	38.61년
Ⅳ등급	49	1,344mm	39.22년



[그림 14] 직접평가 등급과 직경간의 Density Plot

3.2.2 매설년수(Year)

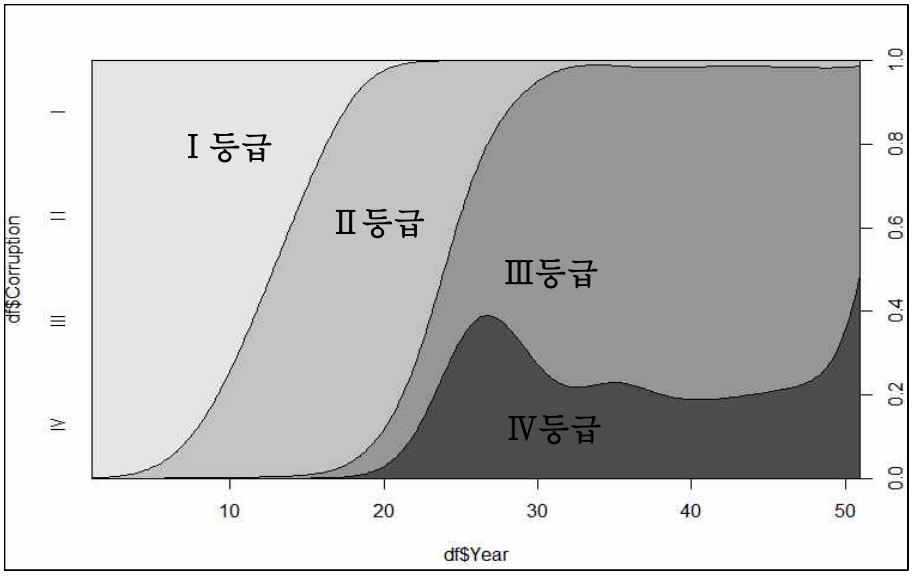
매설년수는 해당관로가 매설된 시점을 기준으로 평가시점까지의 경과년수를 산정하여 적용한다. 매설년수는 관로의 경년화와 노후도를 측정함에 있어 매우 중요한 인자이다. 안홍기(2018)는 관 연령에 따른 부식두께 예측모델을 구축하였다. 관 연령에 따른 부식두께 예측 모델 형태는 기존에 제안된 경험적 모델 형태인 선형함수 모델(Linear model), 거듭제곱 모델(Power model), 지수형 모델(Exponential model)을 이용하였고, 각 모델 내 계수들에 대한 최적 값을 도출하기 위하여 최소 자승법을 이용하였다. 해당 모델에 적용된 주요 변수는 관의 연령 즉, 매설년수로서 상수도 관의 노후도 예측에서 중요한 인자임을 알 수 있다. [18]

구 분	모델형태	계 수	
		기술진단 가이드북	적정개량시기용역
선형함수 모델 (Linear)	kT T = 관 연령 k = 상수	k = 0.057	k = 0.0268
거듭제곱 모델 (Power)	kT^n n = 상수	-	k = 0.0824 n = 0.7888
지수형 모델 (Exponential)	$aT + b(1-e^{-cT})$ a, b, c = 상수	a = 0 b = 1.736 c = 1.737	a = 0.0232 b = 0.3794 c = 1.6890

[그림 15] 부식두께 예측 모델

또한, 분석에 사용된 데이터의 분포를 살펴보면 그림 16과 같이 매설년수가 경과할수록 노후도 등급이 상승하는 것이 명확히 나타난다. 매설 후 10년 이후부터 노후도 평가 Ⅱ등급인 세척이 필요한 관

로가 점차 증가하여 20여년이 경과하면 노후도가 더욱 증가하여 Ⅲ 등급이 많아지는 것을 확인할 수 있다. 20년 이후에는 매설년도 뿐 아니라 매설환경, 수압 등 제반여건에 따라 노후도가 급격히 증가하여 Ⅲ, Ⅳ등급 관로의 비율이 점차 증가함을 알 수 있다. 따라서, 매설년수는 노후도 평가에서 매우 중요한 변수이므로 금회 연구에서 독립변수로 채택한다.



[그림 16] 직접평가 등급과 매설년도간의 Density Plot

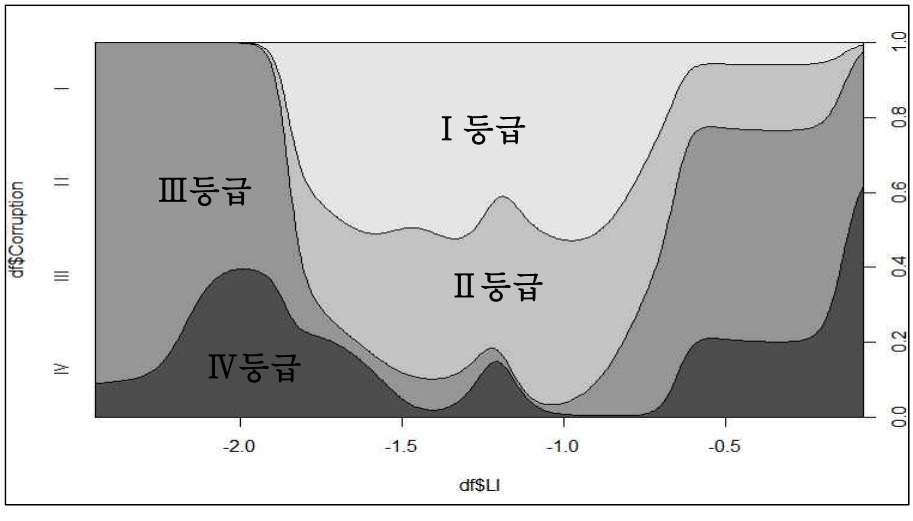
3.2.3 수질부식성(L.I, Langelier Index)

수질부식성은 수돗물이 부식을 유발하는 잠재력을 표현하는 지표이다. 일반적으로 국내에서는 주로 Langelier Index를 통해 수질부식성을 판단한다. LI는 수돗물에서 탄산칼슘의 포화상태로서 부식성의 지표가 된다.[5]

[표 13] LI 수질부식성 지표 산정공식

랑게리아지수(LI) =pHa − pHs 여기서 pHS : 탄산칼슘(CaCO3) 포화 시 수돗물의 pH 수온, pHa, 칼슘이온농도, 알칼리도, 전기전도도 값으로 산출 pHa : 실제 수돗물에서 측정된 pH LI가 0이상이면 스케일이 생성되어 부식이 억제 가능 (non Corrosive) LI가 0 이면 탄산칼슘이 포화상태 (equilibrium) LI가 0 미만이면 관 내부 부식이 발생할 수 있음을 의미 (Corrosive)

수질부식성 지수 측정을 위해서는 정수장 또는 배수지에서 수돗물을 채수하여 pH 등 관련된 수질지표를 측정하여 L.I를 계산한다. 하지만 대부분의 광역상수도는 하나의 취수원에서 취수하여 정수장에서 정수처리 후 공급하기 때문에 같은 광역상수도 시설인 경우 L.I값이 모두 동일하게 적용되는 점, 계절적 특성에 따라 L.I가 달라질 수 있지만 이를 연구 모델에 반영하기가 곤란하고 데이터 수집도 어려운 점, 분석에 사용된 데이터의 분포를 살펴보면 L.I의 변화에 따라 노후도 등급에 미치는 영향을 확인하기 곤란한 점을 고려하여 L.I는 주요 변수로 사용하지 않는 것으로 결정한다.



[그림 17] 직접평가 등급과 수질부식성간의 Density Plot

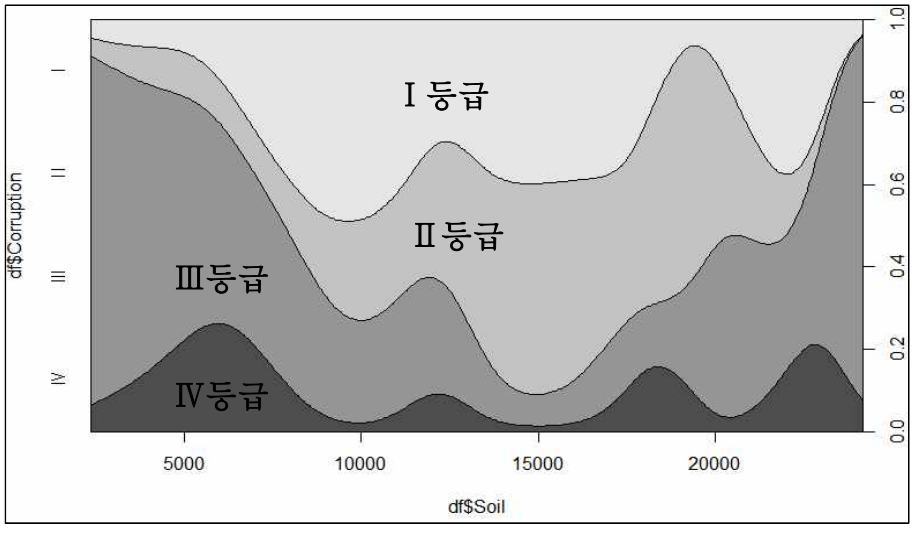
3.2.4 토양 부식성(Soil)

토양부식성은 외부부식에 대한 영향력을 평가하는 지표이다. 토양비저항, 토양 pH, 토양산화환원전위, 토양함수율, 산화물(염화물 또는 황화물)을 주로 측정하여 부식성의 정도를 측정한다. 토양비저항은 토양 내 부식전류의 흐름에 대한 저항력으로 토양비저항이 낮을수록 부식성이 높다는 것을 의미한다. 토양 pH는 토양의 산성도를 나타내는 지표로서 pH 4 또는 pH 8.5이상이면 부식성이 높다는 것을 의미한다. 토양산화환원전위는 토양이 산화적인가 환원적인가를 평가하는 지표로서 산화화원전위가 100mV이상이면 부식성이 낮은 것으로 판단한다. 토양 함수율은 수분함유량이 높을수록 수분 내 용존산소도 많아지므로 철을 산화시켜 부식성을 높일수 있다. 산화물 중 황화합물은 환원성 박테리아에 의한 부식가능성을 보여주므로 미생물에 의한 부식가능성이 있는 것으로 판단한다. 토양부식성은 강관과 직접 접촉되는 토양의 부식특성을 나타내는 지표로서 관로의 부식에 큰 영향을 미치게 된다. 다만, 간접평가에서는 토양부식성을 나타내는 여러 가지 지표 중에서 가장 대표성이 있는 토양비저항 값을 기준으로 토양의 부식성을 평가한다. [4]

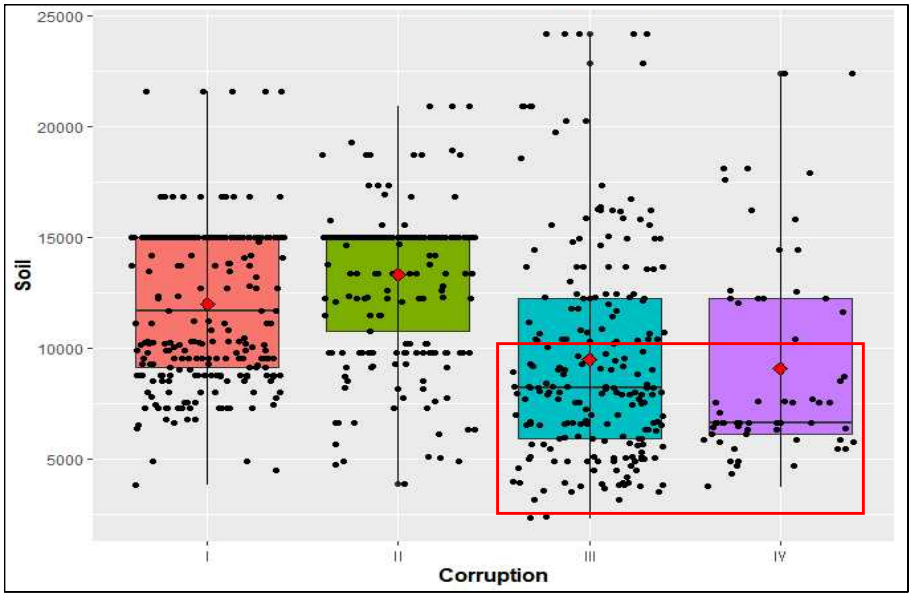
분석에 사용한 데이터의 특성을 살펴보면, Conditional Density에서는 토양비저항값이 노후도 등급에 일관성있는 영향을 미치고 있다고 보기는 어렵다. 하지만, Box plot을 그려 데이터의 분포를 살펴보면 Ⅲ~Ⅳ등급 일수록 토양비저항이 낮은 값이 많은 것을 확인할 수 있으며, 분석과정에서 Ⅲ등급과 Ⅳ등급을 분류함에 있어 중요한 변수로 활용될 수 있을 것이라고 판단하였다.

따라서, 관의 외부부식에 영향이 크고, 실제 관로 주변에서 토양시료

를 채취하여 데이터를 취득하는 점, Ⅲ등급과 Ⅳ등급을 분류시 필요하다고 판단되는 점을 고려하여 금회 연구의 독립변수로 채택하였다.



[그림 18] 직접평가 등급과 토양부식성간의 Density Plot



[그림 19] 직접평가 등급과 토양부식성간의 Box Plot

3.2.5 전기방식(Electronic)

전기방식은 관이 매설된 구간에 미세 전류를 공급하여 전위차를 안정화시킴으로서 관 외부 표면에 부식이 발생되지 않도록 보호하는 방법이다. 광역상수도 시설에서는 일반적으로 희생양극법이나 외부전원법을 사용하고 있다. 고영환(2007)은 '전기방식 적용여부에 따른 관로 부식상태 비교 및 경제성 검토'에서 전기방식 효율 평가 모사시험 조건으로 미방식 조건과 방식조건(희생양극, 외부전원)에서 부식속도를 평가하였다.

실험결과, 미방식 상태에서는 토양비저항이 높은 경우 배관의 부식이 발생하였으나, 유사한 조건에서 전기방식이 된 경우에는 배관의 부식이 발생하지 않는 것으로 실험하였다. 전기방식은 강관의 부식성에 큰 영향을 미치는 인자이며, 실제 전기방식이 설치되지 않은 경우 토양비저항에 의한 관로 부식영향이 매우 커지는 것을 확인할 수 있다. 이에, 현행 간접평가 방식에서는 전기방식의 설치 유무와 전기방식 기준 만족여부를 기준으로 평가하고 있다. [19]

하지만, 2010년 이후에는 전국 모든 광역상수도 강관 시설에 전기방식이 도입되었기 때문에 전기방식 설치 유무는 변수로의 변별력이 없는 점, 전기방식 기준 만족여부는 범주형 변수로서 본 연구에서 중요도가 낮은 점을 고려할 때 전기방식은 변수로 채택하지 않는 것으로 결정하였다.

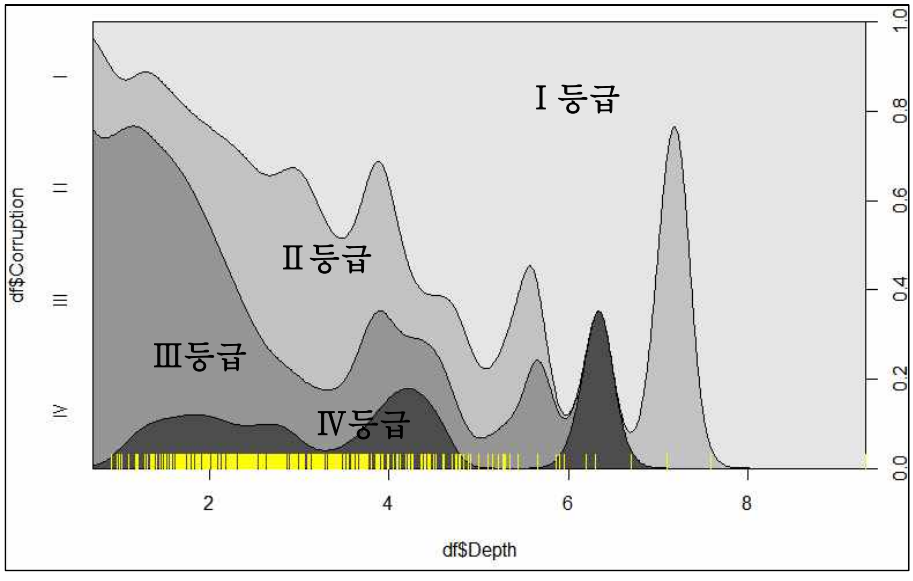
3.2.6 매설깊이(Depth)

매설깊이는 지반고에서 관상단까지의 깊이를 m단위로 적용한 값이다. 매설깊이는 토압에 의한 외부하중을 고려하기 위해 필요한 변수이다. 매설깊이에 따라 토압, 차량하중, 온도하중 등 다양한 외부하중의 영향을 받게 된다. 분석에 사용한 데이터를 살펴보면 매설깊이가 2.5m 이하인 경우에 Ⅲ, Ⅳ등급이 많이 분포하는 것을 확인할 수 있다.

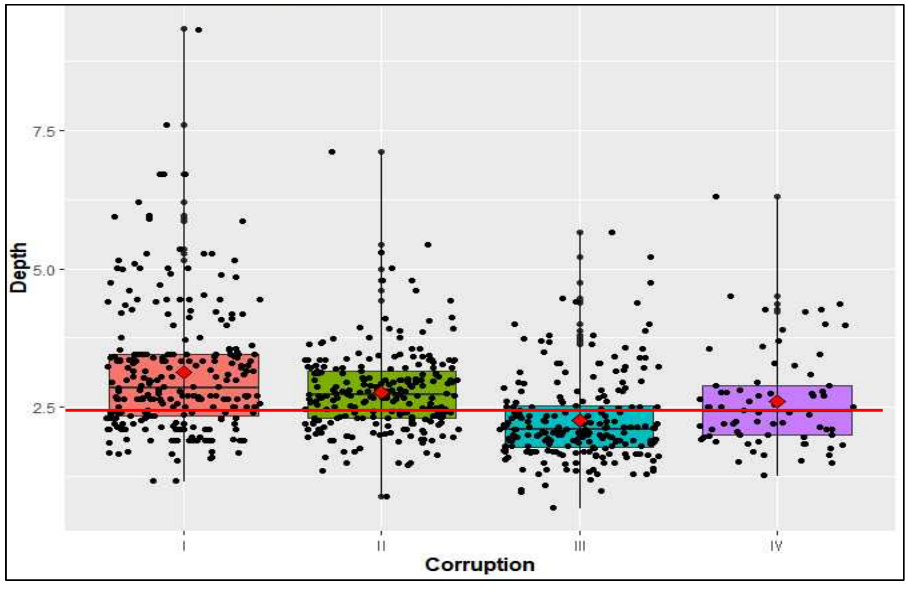
류연선(2017)에 따르면 매설관의 상부 및 인근에 상재하중이 작용할 때 나타나는 매설관의 구조적 거동을 유한요소해석을 통해 분석하였다. 그 결과, 매설관에 발생하는 최대응력은 매설관 상부에 상재하중이 존재하는 구간 내에 발생하였으며, 깊이가 깊어질수록 관에 작용하는 응력이 점차 감소하는 것으로 분석하였다. [20]

김상규(2004, 토질역학 이론과 응용)는 지표면에 놓인 하중으로 생긴 연직응력의 증가는 지표면에 집중 하중이 놓일 때 이로 인한 연직응력의 증가량은 깊이의 제곱에 반비례하고, 하중의 중심에서 멀어질수록 감소한다. 즉, 지중의 연직응력 증가량은 하중의 중심선상에서 가장 크며, 깊이가 깊어짐에 따라 감소하여 어느 깊이 이상에서는 지표면에 있는 하중의 영향을 받지 않게 된다. [21]

따라서, 매설깊이는 관에 작용하는 응력에 영향을 미치게 되며, 응력의 지속적인 작용은 관의 노후도 가속화의 원인이 되는 요인이 되므로 매설깊이는 본 연구의 독립변수로 채택하는 것으로 결정하였다.



[그림 20] 직접평가 등급과 매설깊이간의 Density Plot



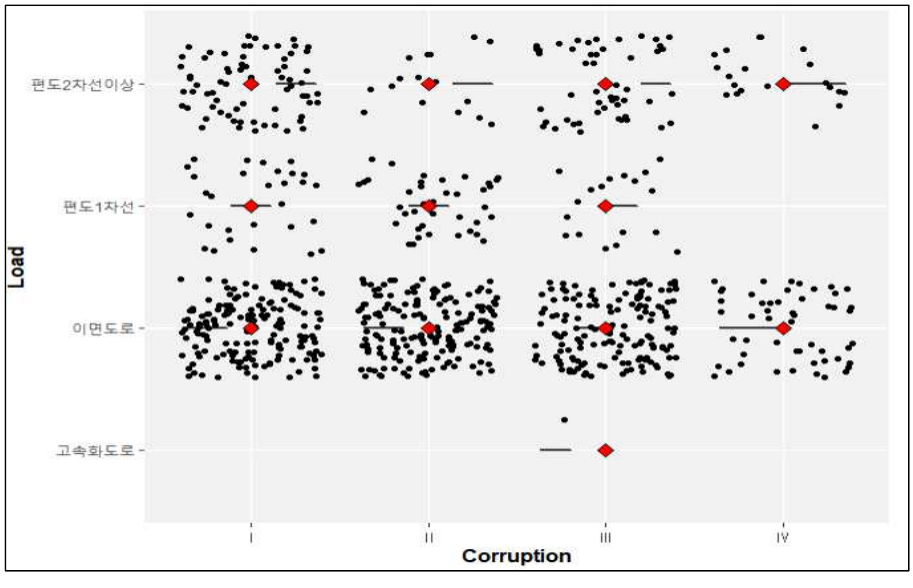
[그림 21] 직접평가 등급과 매설깊이간의 Box Plot

3.2.7 도로형태(Load)

도로형태의 관로의 매설위치에 따라 상부에 작용하는 활하중을 고려하기 위한 변수이다. 관로가 매설된 지역의 도로용도, 편도 차선수에 따라 이면도로(보도, 수도용지), 편도1차선, 편도2차선, 고속화도로로 구분하여 기록한다. 도로형태는 범주형 변수로서 크게 4가지 도로의 종류에 따라 구분하고 있다. 분석에 사용된 데이터의 분포를 살펴보면 노후도 평가등급과 도로형태간의 연관성을 찾기가 어렵다.

따라서, 대부분 이면도로 내에 위치한 관로가 많으며, 데이터의 분포에서 특징을 찾기가 곤란한 점, 변수에 있어 변별력이 낮은 점을 고려하여 독립변수로 채택하지 않는 것으로 결정하였다.

다만, 도로의 형태와 매설깊이, 관경을 종합적으로 고려하여 공식에 적용하면 외부하중에 의한 응력을 계산할 수 있으므로 다음 절에서 변수의 변경시 도로형태를 응용할 예정이다.

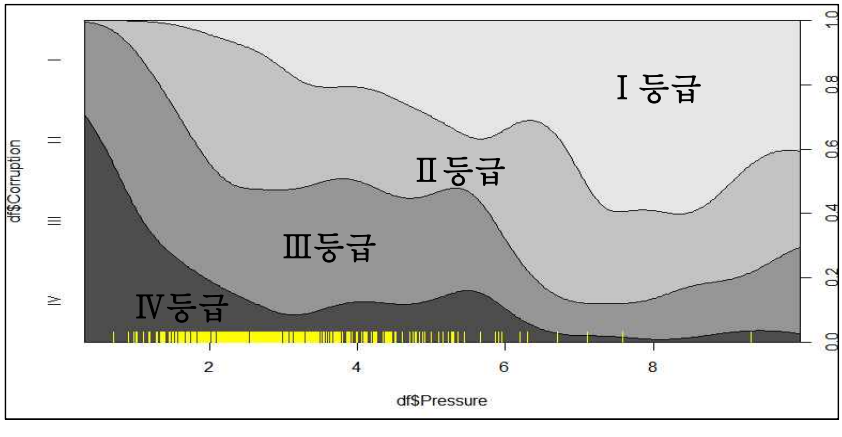


[그림 22] 직접평가 등급과 도로형태 산점도

3.2.8 내부하중(Pressure)

내부하중은 관로에 작용하는 최대수압이다. 표11과 같이 내부수압과 수충격압에 의해 내압에 의한 원주방향 응력이 계산된다. 내압에 의한 응력값과 외압에 의한 응력값을 합산하여 안전계수를 산정하기 때문에 내부하중은 매우 중요한 변수가 된다.

분석에 사용된 데이터의 특성을 살펴보면 내부압력이 낮은 경우엔 노후도 Ⅲ, Ⅳ등급이 많이 분포하는 것을 확인할 수 있다. 이는 에너지 보존의 법칙에 따라 정수두는 동일하다는 가정하에 내부압력이 높아지면 유속이 저하되고, 내부압력이 낮아지면 유속이 빨라지는 베르누이의 정리에 의해 설명할 수 있다. 내부압력이 낮아져서 유속이 빨라지게 되면 관 마찰 손실수두가 증가하게 된다. 관수로에서 유속분포는 관 중심에서 가장 빠르고, 관 벽면에서 가장 느리게 발생한다. 이는 관 벽면에서 관 마찰손실이 발생하기 때문이며, 관 마찰손실 수두는 유속의 제곱에 비례하는데 공식에 따라 마찰손실 수두가 크다는 것은 관 벽면에서의 소류력이 증가하는 것을 의미한다. 이는 관 벽면의 도장재의 손실이나 관체의 손상에 영향이 크다는 것을 알 수 있다. 따라서, 내부하중은 본 연구의 독립변수로 채택한다.



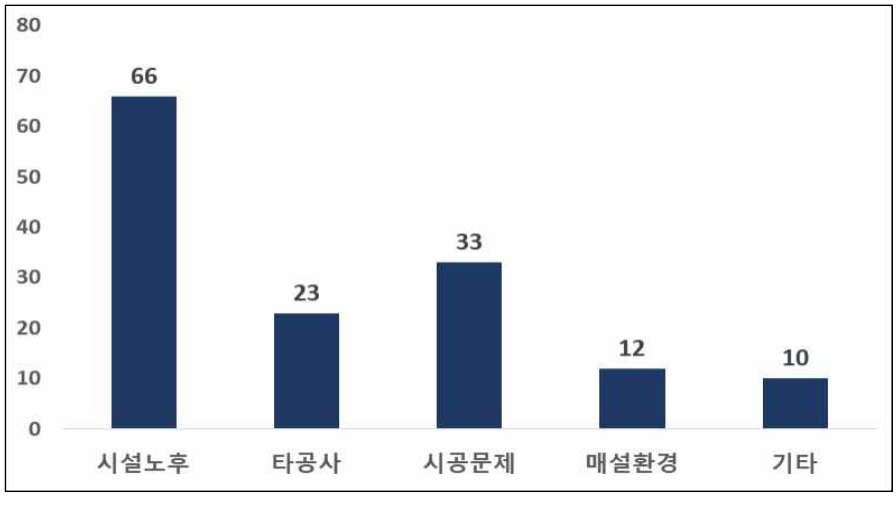
[그림 23] 직접평가 등급과 내부하중간의 Density Plot

3.2.9 사고이력(Accident)

사고이력은 최근 3년간 발생한 관로파손 또는 누수건수를 집계하여 건/km/년으로 환산한 값이다. 파손건수 산정 시 타 공사에 의한 인위적인 파손은 제외하고 자연적인 원인인 노후화, 시공결함으로 인한 관체 손상, 매설환경에 의한 부식 건수로 산정하여야 한다.

하지만, ’15년부터 ‘20년까지 6년간 광역상수도의 누수건수 144건 중 인위적인 파손 23건을 제외하면 자연적인 원인에 의한 파손건수는 121건이다. 이를 연간으로 환산하면 약 20여건이고, 전국 48개 광역상수도 관로 5,681km 구간에 적용하면 매우 희박하게 사고이력이 검출된다.[7]

분석에 사용된 데이터에서 사고이력은 전체 787건 중 21건인 2.6% 구간에서만 확인되는 등 변수로서의 변별력이 낮기 때문에 금회 연구에서는 변수로 채택하지 않는 것으로 결정하였다.



[그림 24] 광역상수도 사고 원인별 현황(’15~‘20)

3.2.10 분기건수(Branch)

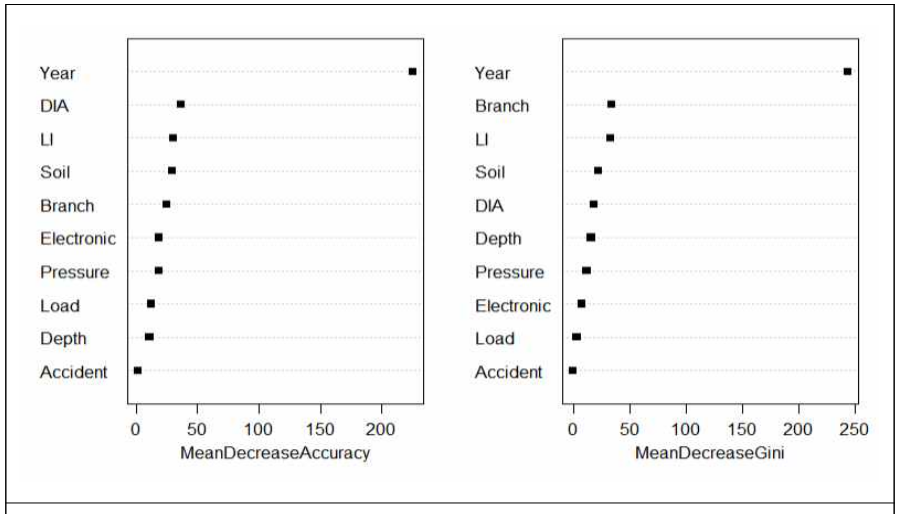
분기건수는 관로상 분기점인 수용가 분기, 각종 밸브실 분기수를 개소수/1km로 환산하여 계산한 값이다. 지자체 상수도의 경우 배수관로에서 급수관로로 연결되는 수많은 분기관로가 있으며, 분기건수가 많으면 재질이 다른 이종관의 접합이나 본관의 천공 등으로 인한 취약지점이 많아져 노후화의 원인이 되기 때문에 분기건수를 간접평가 인자로 사용하고 있다. 하지만 광역상수도는 지자체에 물을 공급하는 시설로서 분기관로의 설치가 매우 적고 제한적이다. 대표적인 분기지점인 밸브실은 건설당시부터 설치하며, 용수공급 분기지점도 대부분 사업계획 단계에서 확정한다.

따라서, 광역상수도는 분기건수가 매우 적어 변수의 변별력이 낮기 때문에 금회 연구에서는 변수로 채택하지 않는 것으로 결정하였다.

3.3 데이터의 특성에 따른 독립변수 채택

앞서 3.2 데이터의 특성에서 간접평가 데이터의 각 변수별 데이터의 특성과 분포를 살펴보았다. 추가로 랜덤포레스트 모델의 변수의 중요도(Importance) 측정 결과를 바탕으로 독립변수를 채택하고자 한다.

그림 25에서 보면 매설년도(Year)가 노후도 평가등급을 결정함에 있어 매우 결정적인 인자이며, 나머지 변수들의 Mean Decease Accuracy (MDA)와 Mean Decrease Gini(MDG) 중요도를 확인 할 수 있다. 이를 바탕으로 3.2에서 검토한 각 변수의 필요성과 상대적으로 중요도를 종합적으로 판단하여 표 14와 같이 최종 분석모델에 활용할 독립변수를 결정하였다.



	I	II	III	IV	MeanDecreaseAccuracy	MeanDecreaseGini
DIA	8.099393	19.841582	13.367168	42.560066	36.442970	18.43300740
Year	221.377512	161.695133	97.137734	86.425885	225.280396	243.79613600
LI	19.192249	28.771977	14.684307	31.493634	30.387154	33.57178074
Soil	5.261003	19.895150	6.135649	29.348407	29.799022	22.39018713
Electronic	8.008048	10.290405	19.049471	7.734729	18.982445	8.01512555
Depth	5.122735	16.407374	-3.133011	14.003409	10.958306	15.95610557
Load	1.223088	6.585912	2.359597	13.620214	12.357125	3.24517119
Pressure	5.647116	13.085544	3.508263	19.894954	18.489291	12.15053994
Accident	0.000000	1.000500	1.414726	0.000000	1.733186	0.04657205
Branch	8.159403	19.922680	17.898396	15.471989	25.132059	34.53159924

[그림 25] 독립변수의 중요도

[표 14] 데이터 특성을 고려한 독립변수 채택 사유

간접평가 항목	데이터 특성상 필요성	Mean Decrease Accracy	Mean Decrease Gini	독립 변수 채택	채택 사유
관경	높음	36.44	18.43	○	필요성과 중요도 높음
매설년도	높음	225.28	243.79	○	필요성과 중요도 높음
수질부식성	낮음	30.38	33.57	×	필요성 낮음, 중요도 높음
토양부식성	보통	29.79	22.39	○	필요성 보통, 중요도 보통
전기방식	낮음	18.98	8.01	×	필요성 낮음, 중요도 낮음
매설깊이	높음	10.95	15.95	○	필요성 높음, 중요도 보통
도로형태	낮음	12.35	3.24	×	필요성과 중요도 낮음
내부하중	높음	18.48	12.15	○	필요성 높음, 중요도 보통
사고이력	낮음	1.73	0.04	×	필요성, 중요도 낮음
분기건수	낮음	25.13	34.53	×	필요성 낮음, 중요도 높음

* 중요도의 판단은 Box plot에서 Q1~Q3 보통, Q1~최대값 높음, 최소값 ~Q3 낮음
(단, 매설년도는 중요도가 매우 높으므로 중요도 판단시 제외하고 결정함)

데이터의 특성상 필요성과 머신러닝 모델 적용 시 중요도를 종합적으로 고려할 때 당초 10가지 간접평가 데이터 중 관경, 매설년도, 토양부식성, 매설깊이, 내부하중 등 5가지 항목을 독립변수로 최종 결정하였다.

3.4 노후도 예측모델의 종속변수 채택

종속변수는 2가지를 채택하였다. 본 연구는 직접조사없이 간접평가 데이터를 활용하여 직접평가를 예측함으로서 노후관 개량사업 시행 시 사업대상 여부와 투자우선순위 결정, 총사업비 산정의 근거로 활용하기 위한 목적이다. 따라서, 종속변수는 노후관 개량사업 대상인지 여부와 직접평가 등급을 예측하는 것으로 하였다.

노후관 개량사업 대상여부 결정은 직접평가 등급 결정 흐름도에 따라 Ⅰ~Ⅱ등급은 계속 사용, 즉 노후관 개량사업 대상이 아닌 것으로 하고, Ⅲ~Ⅴ등급은 노후관 개량사업 대상으로 하였다.

[표 15] 직접평가 등급별 데이터의 수와 개량방법

직접평가 등급	개량방법	데이터 수	종속변수 결정
Ⅰ등급	개량안함	264개	계속사용
Ⅱ등급	세척	233개	
Ⅲ등급	비구조적 갱생	218개	개량대상
Ⅳ등급	구조적 갱생	72개	
Ⅴ등급	교체	없음	분석 제외

노후관 직접평가 등급은 Ⅰ등급에서 Ⅳ등급까지 4가지 등급으로 종속변수를 결정하였다. Ⅴ등급은 안전율이 1이하인 관로로서 이미 파손이나 누수가 발생하여 복구가 완료된 점과 수집데이터 중 Ⅴ등급으로 판정된 관로가 없어 분석이 곤란한 점을 고려하여 종속변수에서 제외하였다.

제 4 장 강관의 노후도 예측 모델

4.1 분석방법 결정

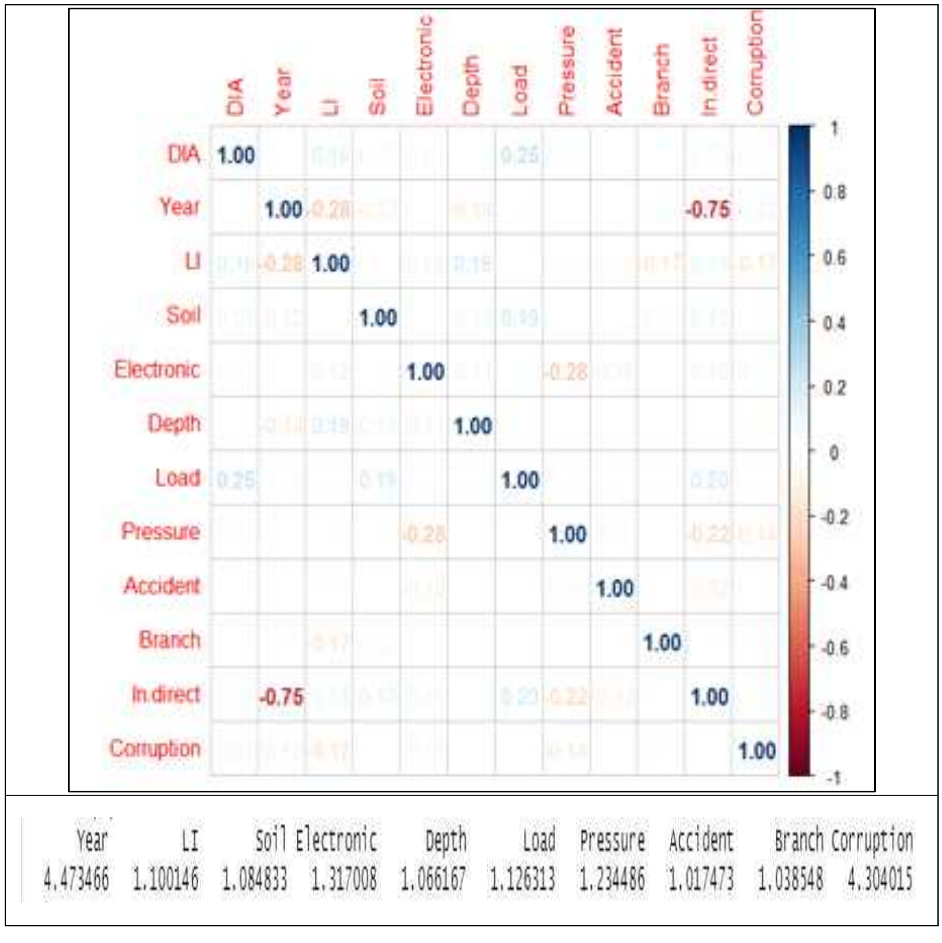
머신러닝 기법으로 상수도 강관의 노후도 등급 예측을 위해서 회귀분석(Regression)과 분류분석(Classification)방법을 검토하였다. 이윤호(2020)는 랜덤포레스트 기법을 이용한 상수관로 부식에 관한 연구에서 수도권광역상수도 Ⅲ단계 시설에서 취득한 데이터 283개를 활용하여 부식깊이 비율을 직접 예측하고, 상수관망 정밀조사 매뉴얼에서 제시한 부식깊이 비율 산정 모델과 비교하였다. 해당 연구는 회귀분석 방법을 이용하였으며, 연구결과, 실제 부식깊이 비율 0.06~0.19 대비 0.07~0.16으로 실제 부식깊이 비율과 유사하게 예측하였다. [22]

이윤호(2020)의 연구는 단일 광역상수도 시설에서 취득한 데이터를 활용하여 간접평가 자료를 독립변수로 하고, 부식깊이 비율 값을 직접 예측하는 모형이다. 본 연구는 전국 11개 광역 및 공업용수도 시설에서 취득한 데이터 787개를 활용하여 같은 방법으로 회귀분석을 실시하였다.

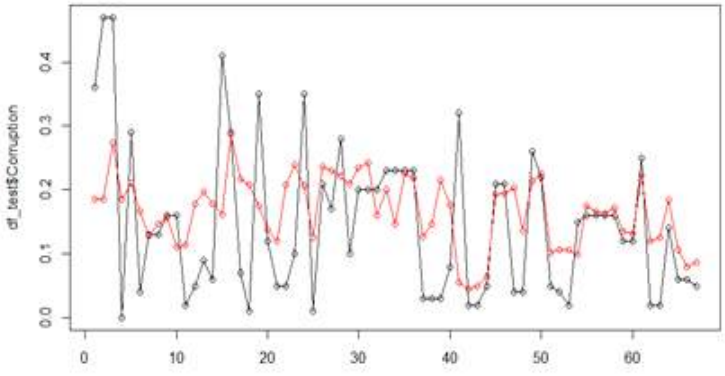
먼저 회귀분석을 위해 각 독립변수(간접평가 10가지 항목)와 종속변수(부식깊이 비율)의 상관관계 Matrix를 그려 살펴보았다. 그 결과, 독립변수와 종속변수의 상관관계가 0.02~-0.17로 낮으며, 랜덤포레스트를 이용하여 부식깊이 비율을 예측한 결과, 독립변수의 설명력이 23%로 낮고, RMSE가 0.21로 모델의 성능이 낮은 것으로 확인하였다. 또한, 변수간의 다중공선성은 1.06~4.47로 다중공선성 판단기준인 9보다 낮아 다중공선성 문제는 발생하지 않는 것으로 확인되었다.

선행연구와 비교할 때, 본 연구는 전국 데이터를 활용한 점, 변수간의 상관관계가 낮은 점, 선형연구 데이터는 단수조사의 한계로 인해 부득이 인접 지점의 직접조사 결과를 사용한 데이터가 있는 점에서 차이가 있다.

이에, 회귀분석 방법으로는 분석이 곤란할 것으로 판단되어 분류분석(Classification) 방법을 적용하는 것으로 결정하였다.



[그림 26] 변수간 상관관계 행렬 및 다중공선성



[그림 27] 회귀분석 결과

4.2 분석 알고리즘 및 성능평가 방법

일반적으로 분류분석을 위한 기계 학습 알고리즘은 의사결정 트리, 인공신경망, 서포트 벡터 머신과 앙상블 기법인 랜덤포레스트, XGBoost 기법이 대표적이다. 특히, 앙상블 기법은 트리 기반의 기계 학습 방법으로 다수의 트리 모델을 만들고, 그 결과를 결합하여 Voting 등의 방법으로 예측하기 때문에 하나의 모델을 사용할 때보다 과적합의 영향을 감소시킬 수 있다.

이에 본 연구에서는 분류분석을 위해 앙상블 기법인 랜덤포레스트(Random Forest), XGBoost를 적용하고, 서포트벡터머신(SVM) 등 3가지 머신러닝 알고리즘을 사용하였다.

먼저, 랜덤포레스트(Random Forest)는 Brieman(2001)에 의해 개발되었으며, 의사결정나무 분석의 한계를 극복하기 위해 발전된 심화 기법으로 각 데이터를 한 개의 나무가 아니라 여러 개의 의사결정나무로 확장시켜 분석하는 머신러닝 기법이다. 이는 자유로운 변수 선택을 통해 데이터셋이 과적합(Overfitting)되는 것을 방지하여 보다

높은 정확도를 가진 모형을 구축할 수 있는 앙상블 기법이다.

서포트벡터머신(support vector machine, SVM)은 주어진 샘플 그룹에 대해 분류 규칙을 찾아내는 기법 중 하나이다. 1995년 러시아의 수학자 블라디미르 바프닉(Valdimir N. Vapnik, 1936-)이 처음 제시하였다. 대표적인 기계 학습 알고리즘으로 다양한 분류 문제에 활용되는 알고리즘이다

다음으로 XGBoost 기법은 eXtreme Gredient Boosting의 약자로서 그래디언트 부스팅 방법의 일종이다. 다수의 Decision Tree를 만들어 첫 번째 Tree에서 예측이 잘못된 값에 가중치를 반영하여 다음 모델에 반영한다. 이와 같은 방식으로 학습을 반복하는 병렬학습을 통해 예측의 오류를 최소화하는 머신러닝 알고리즘이다.

위 3가지 분석 알고리즘을 각 모델에 적용하여 강관의 노후도를 예측하면서 모델의 성능을 분석하여 금회 연구에 가장 적합한 알고리즘을 찾는 방식으로 연구를 수행하였다.

머신러닝을 위해 전체 787개 데이터 항목 중 Training 데이터는 70%인 551개와 Test 데이터 30%로 분할하여 연구를 진행하였다.

머신러닝 분류 모델의 성과 검증은 오분류표(Confusion Matrix)와 ROC를 이용한 AUC로 수행하였다. 오분류표는 예측값과 실제값을 비교하여 Positive Value와 Negative Value의 예측력을 평가한다. 이를 통해 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1 score (정밀도와 재현율의 조화평균)을 산정할 수 있다.

AUC(Area Under Curve)는 ROC의 곡선아래에 있는 면적을 의미하며, y축에 민감도, x축에는 1-specificity로 그려, 클래스를 구별하는 모델의 성능을 평가하는데 활용하였다.

[표 16] 분류모델 성능검증을 위한 오분류표

Confusion Matrix		예측값 (Predicted)		
		True	False	평가
Actual (실제값)	True	True Positive (TP)	False Negative (FN)	Sensitivity =Recall (TP/(TP+FN)
	False	False Positive (FP)	True Negative (TN)	Specificity (TN/(TN+FP)
	평가	Precision = Positive Predictive Value (TP/(TP+FP)	Negative Predictive Value (TN/(TN+FN)	Accuracy (TP+TN)/ (TP+TN+FP+ FN)

분석을 위한 각 알고리즘별 파라미터 세팅은 다음과 같다.

랜덤포레스트 모델은 하이퍼 파라미터를 Defalut값을 적용하였으며, 의사결정나무의 수는 여러차례 시행착오를 통해 2,000으로 적용하였다. 무작위로 선택되는 변수의 개수(mtry)는 3으로 설정하였으며, 분석 결과 최적트리수는 189개로 확인되었다. SVM 모델은 SVM Type을 C-Classification으로 SVM-Kernel은 radial, Support vector의 수는 338, COST 1, gamma는 'scale'로 대부분 Defalut값을 적용하였다. XGBoost 모델은 max_depth 5, gamma 0, eta 0.3, min-child-weight 1로 파라미터를 세팅하였다.

사용한 3가지 알고리즘 모두 파라미터는 Defalut값 사용을 우선하였는데, 현재는 데이터의 수가 적어 초기 모델을 구축하는 것이 중요하다고 판단하였다. 향후 충분한 데이터가 확보된 이후 하이퍼 파라미터 세팅을 통해 예측모델의 고도화가 필요하다.

4.3 독립변수 채택의 적정성 검증

본 연구에서 독립변수는 간접평가 데이터 10가지 중 필요성과 중요도가 낮은 5가지 항목을 제외한 관경, 매설년도, 토양부식성, 매설깊이, 내부하중 등 5가지 항목을 독립변수로 최종 결정하였다. 이에, 독립변수 채택이 적정하였는지 검증하였다. 이를 위해 간접평가 데이터 10가지를 모두 독립변수로 채택하였을 경우(Case 1)와 독립변수를 5가지로 채택하였을 경우(Case 2)를 비교하여 모델의 성능이 향상되었는지를 검증하였다.

4.3.1 노후관 개량대상 여부에 대한 예측 결과

노후관 개량대상 여부에 대한 예측결과를 보면 모든 모델에서 Case 1과 Case2가 거의 비슷한 결과를 갖는 것으로 확인하였다. AUC를 기준으로 살펴보면 랜덤포레스트 모델은 1.2%가량 하락, SVM 모델은 1.6%가량 향상, XGBoost 모델은 0.1% 하락하여 독립변수가 많은 Case 1이 모델의 정확도 향상에 영향을 미치지 않는 것으로 확인하였다.

[표 17] Case별 노후관 개량대상여부 예측 성능 비교

구 분		랜덤포레스트	SVM	XGBoost
Accuracy	Case 1	98.73%	95.34%	97.46%
	Case 2	97.89%	97.03%	97.03%
	차이 (2-1)	-0.84%	+1.69%	-0.43%
AUC	Case 1	98.80%	95.60%	96.80%
	Case 2	97.60%	97.20%	97.80%
	차이 (2-1)	-1.20%	+1.60%	+1.00%

4.3.2 직접평가 등급에 대한 예측 결과

직접평가 등급에 대한 예측결과를 보면 Accuracy는 모든 모델에서 같은 값으로 확인되었으며, AUC를 기준으로 보면 독립변수를 소거한 Case2의 경우가 약간의 성능향상이 있는 것으로 확인되었다. 즉, 독립변수의 수가 많다고 하여 모델의 성능이 향상되지 않았으며, 설명력이 낮은 변수는 오히려 불순도를 증가시켜 모델의 성능을 저하시키는 것으로 확인되었다.

[표 18] Case별 직접평가 등급 예측 성능 비교

구 분		랜덤포레스트	SVM	XGBoost
Accuracy	Case 1	92.83%	83.33%	92.74%
	Case 2	92.83%	83.33%	92.74%
	차이 (2-1)	-	-	-
AUC	Case 1	97.08%	86.20%	93.28%
	Case 2	97.54%	91.50%	93.48%
	차이 (2-1)	+0.50%	+5.30%	+0.20%

각 Class별 Sensitivity를 보면 랜덤포레스트 모델을 기준으로 Ⅲ등급 분류성능은 소폭 하락(Case1 91.04% → Case2 88.06%, -2.98%)하였으나, Ⅳ등급의 분류성능은 향상(Case1 50.0% → Case2 61.9%, +11.9%)된 것을 확인할 수 있다. 본 연구 모델에서는 노후관 개량대상 중 Ⅲ, Ⅳ등급을 잘 분류하는 것이 연구목적을 달성함에 있어 매우 중요한 부분이므로 독립변수를 제거하여 채택한 것이 적절하였다는 것을 검증할 수 있다.

```
Overall Statistics

               Accuracy : 0.9283
                 95% CI : (0.8876, 0.9577)
    No Information Rate : 0.3333
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.8989

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: I Class: II Class: III Class: IV
Sensitivity            1.0000    1.0000    0.9104   0.50000
Specificity            1.0000    0.9810    0.9412   0.98140
Pos Pred Value         1.0000    0.9634    0.8592   0.73333
Neg Pred Value         1.0000    1.0000    0.9639   0.95045
Prevalence             0.2911    0.3333    0.2827   0.09283
Detection Rate         0.2911    0.3333    0.2574   0.04641
Detection Prevalence   0.2911    0.3460    0.2996   0.06329
Balanced Accuracy      1.0000    0.9905    0.9258   0.74070
```

[그림 28] Case 1의 Overall Statistics

```
Overall Statistics

               Accuracy : 0.9283
                 95% CI : (0.8876, 0.9577)
    No Information Rate : 0.3333
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.8992

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: I Class: II Class: III Class: IV
Sensitivity            1.0000    0.9857    0.8806   0.61905
Specificity            1.0000    0.9760    0.9529   0.97685
Pos Pred Value         1.0000    0.9452    0.8806   0.72222
Neg Pred Value         1.0000    0.9939    0.9529   0.96347
Prevalence             0.3333    0.2954    0.2827   0.08861
Detection Rate         0.3333    0.2911    0.2489   0.05485
Detection Prevalence   0.3333    0.3080    0.2827   0.07595
Balanced Accuracy      1.0000    0.9809    0.9168   0.79795
```

[그림 29] Case 2의 Overall Statistics

4.4 노후관 개량사업 대상여부 예측 모델

본 절에서는 노후관 개량사업 대상여부를 종속변수로 하는 노후도 예측모델을 3가지 알고리즘을 활용하여 분석하였다.

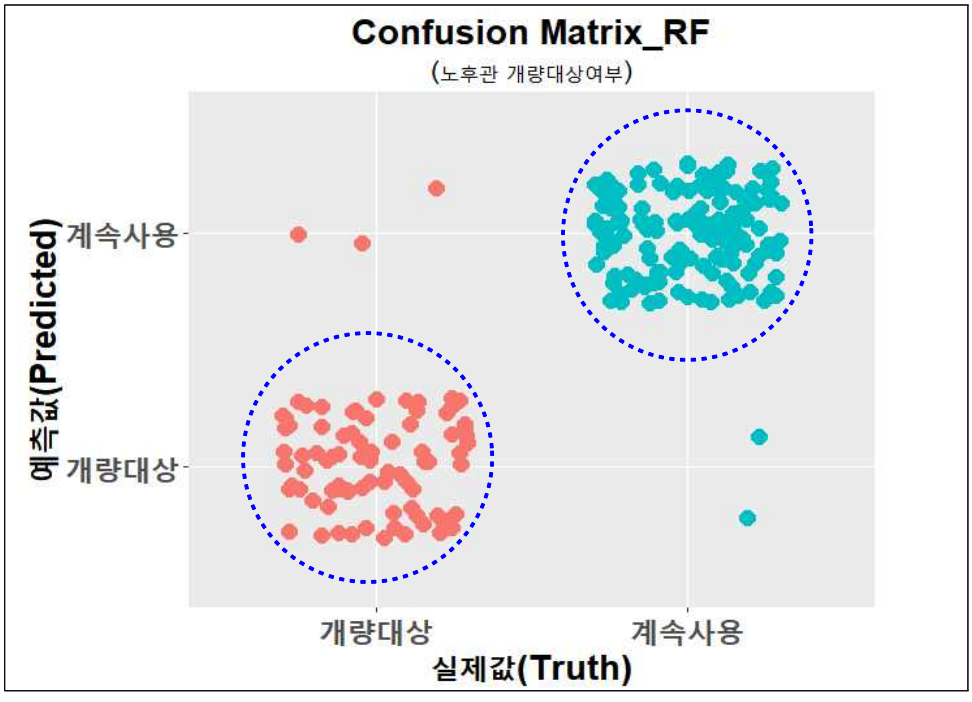
4.4.1 랜덤포레스트를 이용한 노후관 개량사업 대상여부 예측

랜덤포레스트를 이용한 노후관 개량사업 대상여부 예측에서는 높은 정확도를 가지는 예측모델을 구현하였다. Confusion Matrix를 이용하여 오차율을 살펴보면, 모델의 Positive Class는 개량대상이며, 테스트 데이터 237개 중 5건을 제외한 나머지 232건의 예측값이 실제와 일치하였다. 또한, Training 데이터의 OOB estimate of error rate은 1.64%로 Test 데이터의 error rate과 유사한 수준으로 모델이 과적합 되지 않았음을 확인할 수 있다.

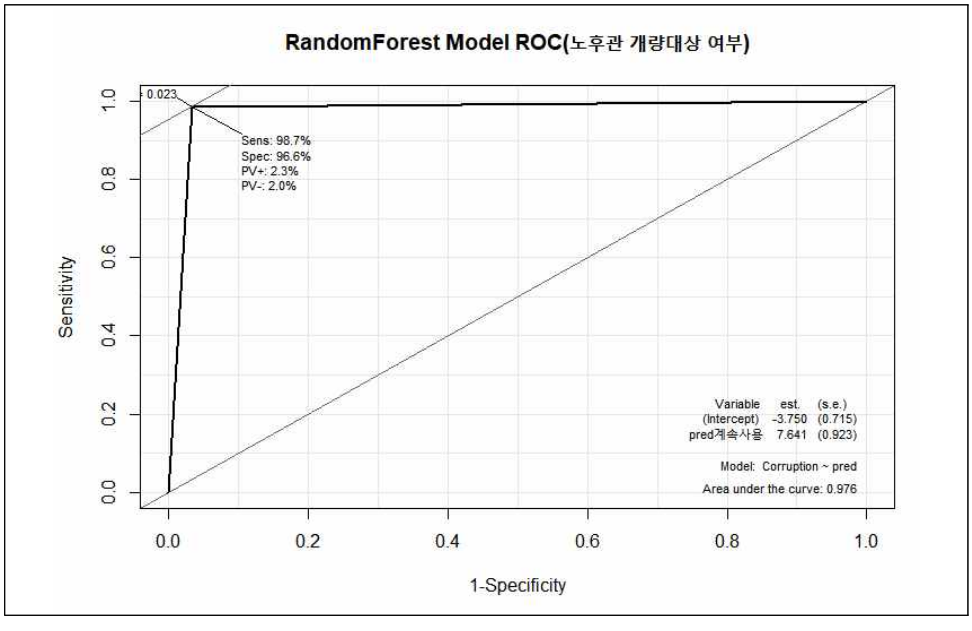
모델의 성능검증을 위해 Confusion Matrix와 ROC를 이용한 AUC를 산정하였다. 랜덤포레스트를 이용한 예측 모델의 Accuracy는 97.89%, AUC 97.60%로 정확도 있는 사업대상 분류 모델임을 확인할 수 있다.

```
Confusion Matrix and Statistics

          Reference
Prediction 개량대상 계속사용
  개량대상       85        2
  계속사용        3      147

               Accuracy : 0.9789
                 95% CI : (0.9515, 0.9931)
    No Information Rate : 0.6287
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.9547

 Mcnemar's Test P-Value : 1

            Sensitivity : 0.9659
            Specificity : 0.9866
         Pos Pred Value : 0.9770
         Neg Pred Value : 0.9800
             Prevalence : 0.3713
         Detection Rate : 0.3586
   Detection Prevalence : 0.3671
      Balanced Accuracy : 0.9762

       'Positive' Class : 개량대상
```

[그림 30] Confusion Matrix(RF_사업대상여부)



[그림 31] Confusion Matrix2(RF_사업대상여부)



[그림32] ROC (RF_사업대상여부)

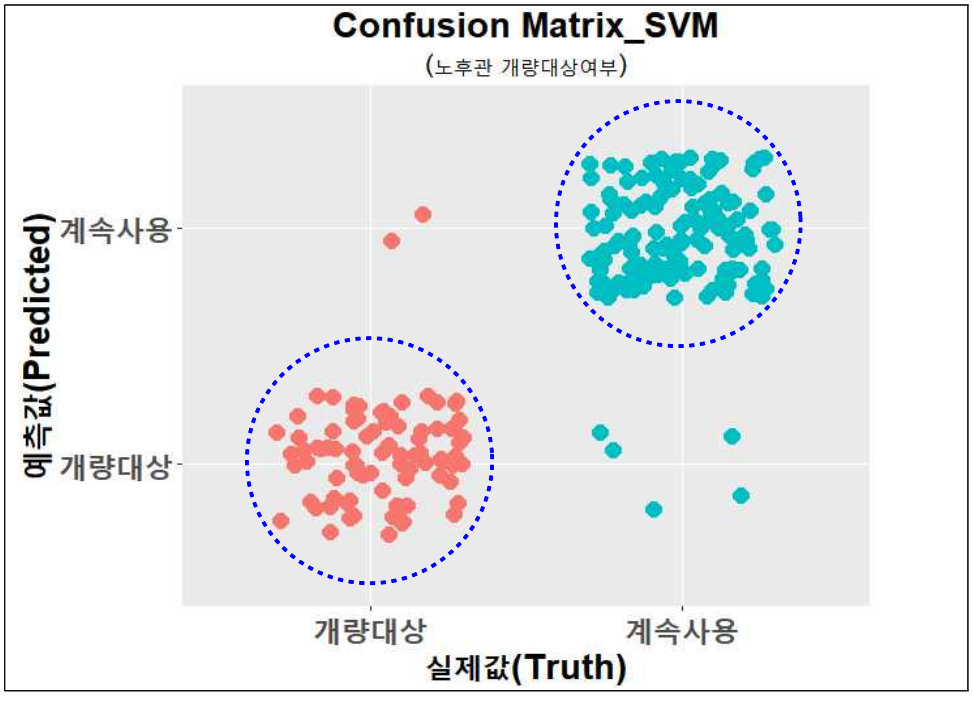
4.4.2 SVM을 이용한 노후관 개량사업 대상여부 예측

서포트 벡터 머신을 이용한 노후관 개량사업 대상여부 예측 모델도 랜덤포레스트 모델과 유사한 수준의 정확도를 확인할 수 있다.

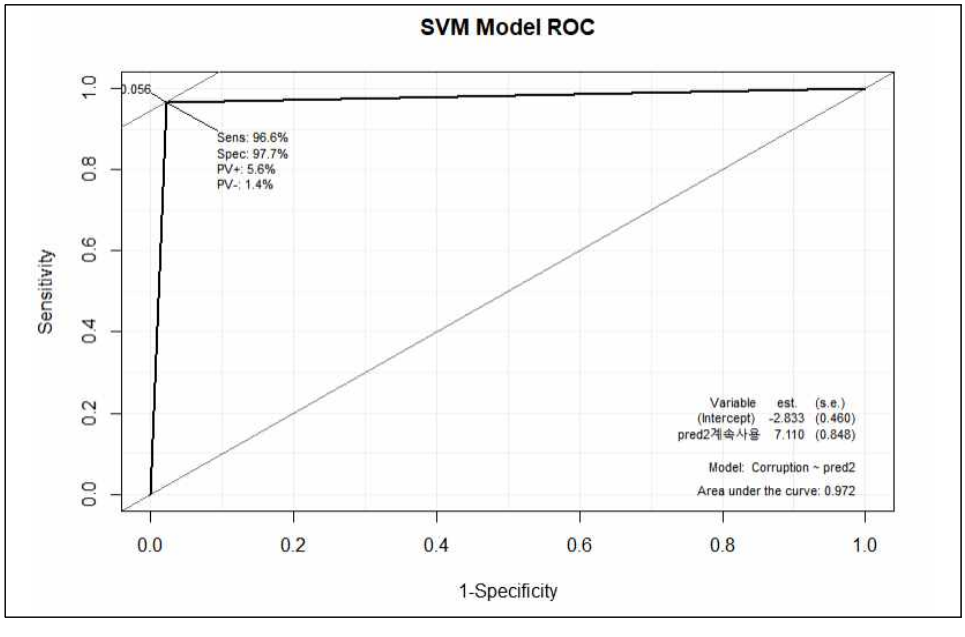
모델의 성능검증을 위해 Confusion Matrix와 ROC를 이용한 AUC를 산정하였다. 서포트 벡터 머신을 이용한 예측 모델의 Accuracy는 97.03%, AUC 97.20%로 정확도 있는 사업대상 분류 모델임을 확인할 수 있다. 본 연구에서 개량대상과 계속사용여부를 구분하는 가장 중요도가 높은 변수는 매설년수인데, 매설년수가 오래될수록 노후관 개량사업 대상이 될 가능성이 매우 높기 때문에 노후관 개량사업 대상여부를 예측하는 것은 어떤 모델에서나 높은 정확도를 보이고 있다.

```
Confusion Matrix and Statistics

          Reference
Prediction 개량대상 계속사용
  개량대상       85        5
  계속사용        2      144

               Accuracy : 0.9703
                 95% CI : (0.9398, 0.988)
    No Information Rate : 0.6314
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.9367

 Mcnemar's Test P-Value : 0.4497

            Sensitivity : 0.9770
            Specificity : 0.9664
         Pos Pred Value : 0.9444
         Neg Pred Value : 0.9863
             Prevalence : 0.3686
         Detection Rate : 0.3602
   Detection Prevalence : 0.3814
      Balanced Accuracy : 0.9717

       'Positive' Class : 개량대상
```

[그림 33] Confusion Matrix(SVM_사업대상여부)



[그림 34] Confusion Matrix2(SVM_사업대상여부)



[그림35] ROC (SVM_사업대상여부)

4.4.3 XGBoost를 이용한 노후관 개량사업 대상여부 예측

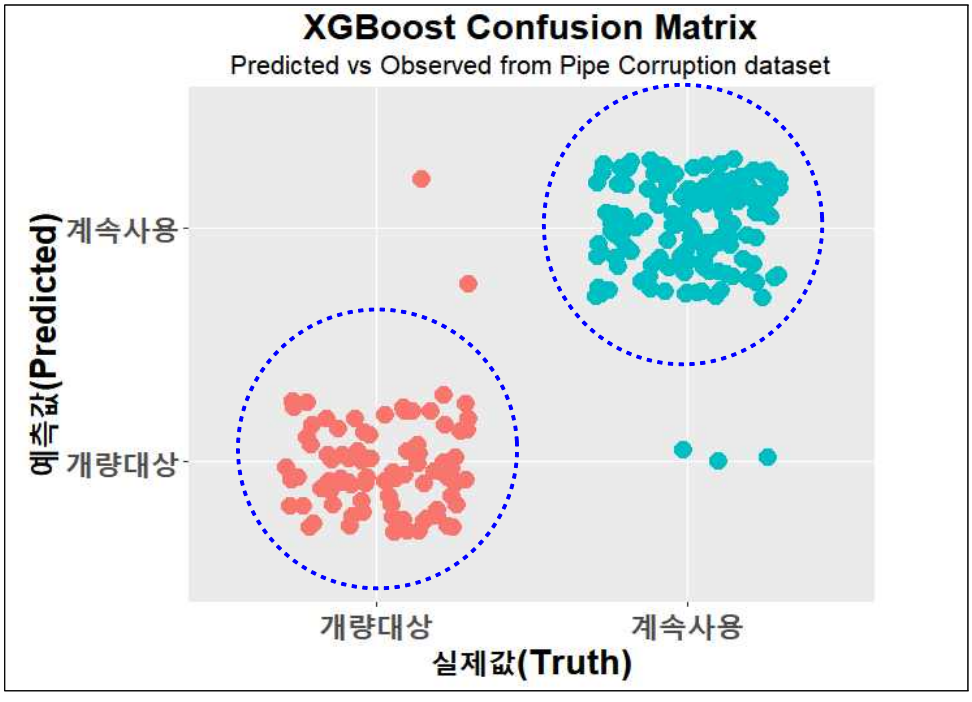
XGBoost를 이용한 노후관 개량사업 대상여부 예측 모델도 다른 예측 모델과 유사한 수준의 정확도를 확인할 수 있다.

모델의 성능검증을 위해 Confusion Matrix와 ROC를 이용한 AUC를 산정하였다. XGBoost를 이용한 예측 모델의 Accuracy는 97.88%, AUC 97.80%로 정확도 있는 사업대상 분류 모델임을 확인할 수 있다.

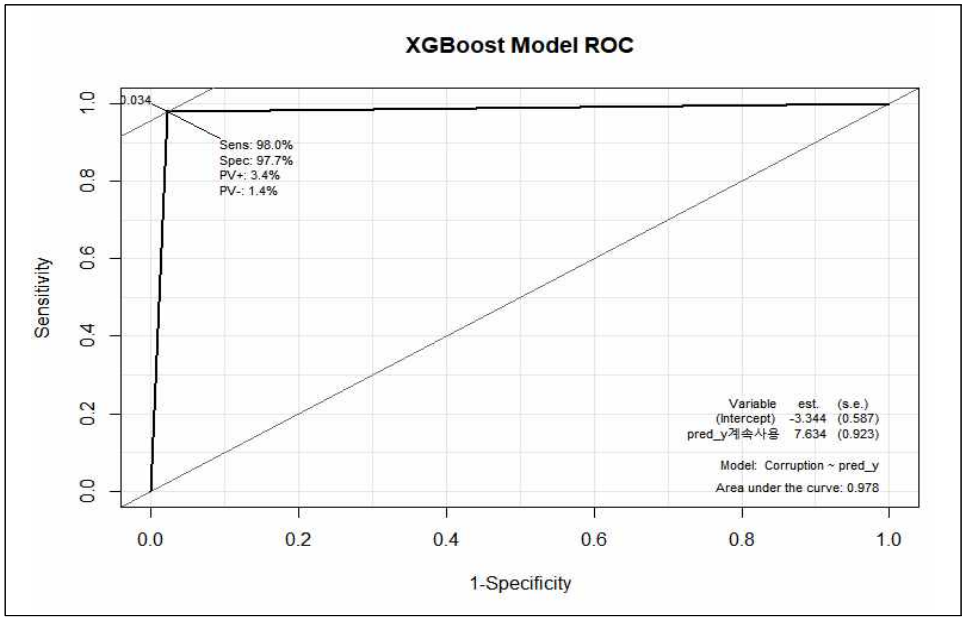
Posivive Class인 개량대상과 Negative Class인 계속사용 모든 항목에서 고르게 97% 수준의 정확도를 확인할 수 있다. 다만, 모든 모델에서 개량대상임에도 계속사용으로 예측한 경우는 적기 개량시기를 놓칠 수 있는 문제점이 있어 개선이 필요하다.

```
Confusion Matrix and Statistics

          Reference
Prediction 개량대상 계속사용
  개량대상       85        2
  계속사용        3      146

               Accuracy : 0.9788
                 95% CI : (0.9513, 0.9931)
    No Information Rate : 0.6271
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.9546

 Mcnemar's Test P-Value : 1

            Sensitivity : 0.9659
            Specificity : 0.9865
         Pos Pred Value : 0.9770
         Neg Pred Value : 0.9799
             Prevalence : 0.3729
         Detection Rate : 0.3602
   Detection Prevalence : 0.3686
      Balanced Accuracy : 0.9762

       'Positive' Class : 개량대상
```

[그림 36] Confusion Matrix(XGBoost_사업대상여부)



[그림 37] Confusion Matrix2(XGBoost_사업대상여부)



[그림38] ROC (XGBoost_사업대상여부)

4.4.4 노후관 개량사업 대상여부 예측 결과

3가지 알고리즘을 이용하여 노후관 개량사업 대상여부를 예측한 결과, 모든 모델에서 약 97% 정도의 정확도를 확인할 수 있었다. 노후관 개량사업 대상여부 예측을 통해 노후도 등급을 추정하기 이전에 1차적으로 대분류가 가능하며, 모델의 정확도를 고려할 때 실제 사업에서도 적용가능한 수준임을 확인할 수 있다.

[표 19] 분류 알고리즘별 노후관 개량사업 대상여부 예측 성능

구 분		랜덤포레스트	SVM	XGBoost
모델의 성능	Accuracy	97.89%	97.03%	97.88%
	AUC	97.60%	97.20%	97.80%

노후관 개량사업 대상여부를 예측하면 계속 사용의 대상이 되는 Ⅰ, Ⅱ등급 그룹과 개량 대상이 되는 Ⅲ, Ⅳ등급 그룹으로 구분을 할 수 있다.

4.5 직접평가 등급 예측에서는 Ⅰ등급에서 Ⅳ등급까지의 4가지 등급에 대한 Multi-Class 분류분석을 실시하여 직접평가 등급을 분류하고, 각 등급별로 개량방법을 결정하는데 활용하고자 한다.

4.5 직접평가 등급 예측 모델

본 절에서는 직접평가 등급을 종속변수로 하는 노후도 예측모델을 3가지 알고리즘을 활용하여 분석하였다.

4.5.1 랜덤포레스트를 이용한 직접평가 등급 예측

랜덤포레스트를 이용하여 직접평가 등급을 예측하였다. Confusion Matrix를 이용하여 Traning 데이터의 OOB estimate of error rate은 5.45%로 Traning 데이터 550건 중 520건에서 실제값과 예측값이 일치하는 것을 확인할 수 있다. 각 Class별로 살펴보면 Ⅰ등급은 에러율이 없고, Ⅱ등급은 1.84%, Ⅲ등급은 5.29%으로 비교적 양호한 에러율을 보이고 있다. 하지만, Ⅳ등급은 37.2%의 에러율을 보이고 있으며, Ⅳ등급으로 예측하였으나, 실제값은 Ⅲ등급인 경우가 19건으로 Ⅲ등급과 Ⅳ등급을 구분하기 위해서는 보다 개선된 연구가 필요한 것으로 판단된다.

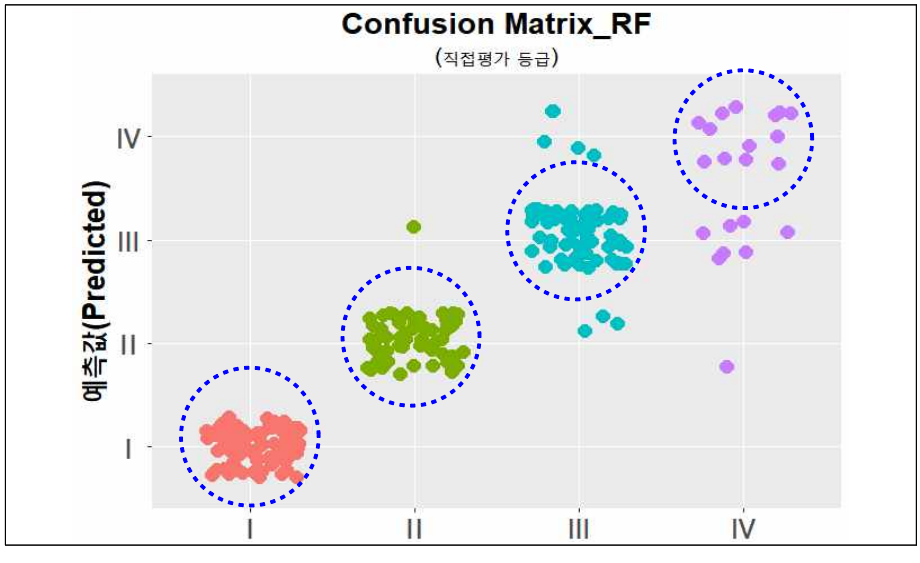
```
Call:
 randomForest(formula = df_train$Corruption ~ ., data = df_train,
                Type of random forest: classification
                      Number of trees: 1000
No. of variables tried at each split: 5

        OOB estimate of  error rate: 5.45%
Confusion matrix:
     I  II  III IV class.error
I 185   0    0  0  0.00000000
II    0 160    3  0  0.01840491
III   0   2 143  6  0.05298013
IV    0   0  19 32  0.37254902
```

[그림39] OOB error rate(RF_직접평가 등급)

모델의 성능검증을 위해 Confusion Matrix와 ROC를 이용한 AUC를 산정하였다. 랜덤포레스트를 이용한 예측 모델의 Accuracy는 92.83%, AUC 97.54%로 종합적인 성능은 개선된 것을 확인할 수 있다. Ⅰ등급과 Ⅱ등급은 거의 100%에 이르는 높은 분류성능이 있지만, Ⅲ등급은 88.01%, 특히, Ⅳ등급의 Sensitivity가 61.9%로 실무에 적용하기 위해서는 해당 Class의 개선이 필요함을 알 수 있다.

```
Confusion Matrix and Statistics

          Reference
Prediction I Ⅱ Ⅲ Ⅳ
        I 79  0  0  0
        Ⅱ  0 69  3  1
        Ⅲ  0  1 59  7
        Ⅳ  0  0  5 13

Overall Statistics

               Accuracy : 0.9283
                 95% CI : (0.8876, 0.9577)
    No Information Rate : 0.3333
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.8992

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: I Class: Ⅱ Class: Ⅲ Class: Ⅳ
Sensitivity            1.0000   0.9857   0.8806   0.61905
Specificity            1.0000   0.9760   0.9529   0.97685
Pos Pred Value         1.0000   0.9452   0.8806   0.72222
Neg Pred Value         1.0000   0.9939   0.9529   0.96347
Prevalence             0.3333   0.2954   0.2827   0.08861
Detection Rate         0.3333   0.2911   0.2489   0.05485
Detection Prevalence   0.3333   0.3080   0.2827   0.07595
Balanced Accuracy      1.0000   0.9809   0.9168   0.79795
```

[그림40] Confusion Matrix(RF_직접평가 등급)



[그림41] Confusion Matrix2(RF_직접평가 등급)

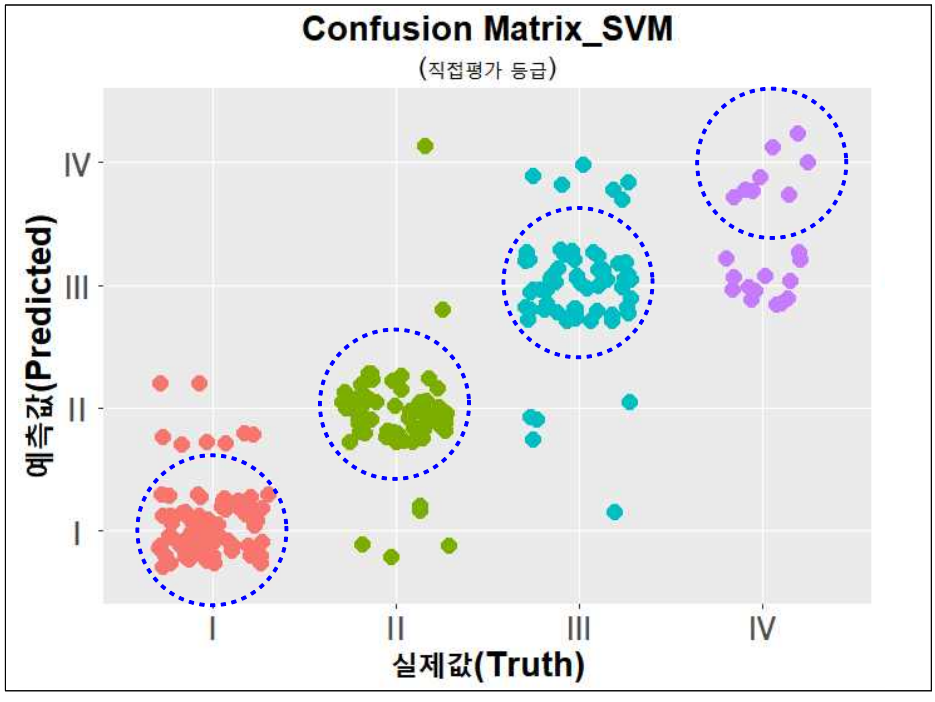
4.5.2 SVM을 이용한 직접평가 등급 예측

서포트 벡터 머신을 이용하여 직접평가 등급을 예측하였다.

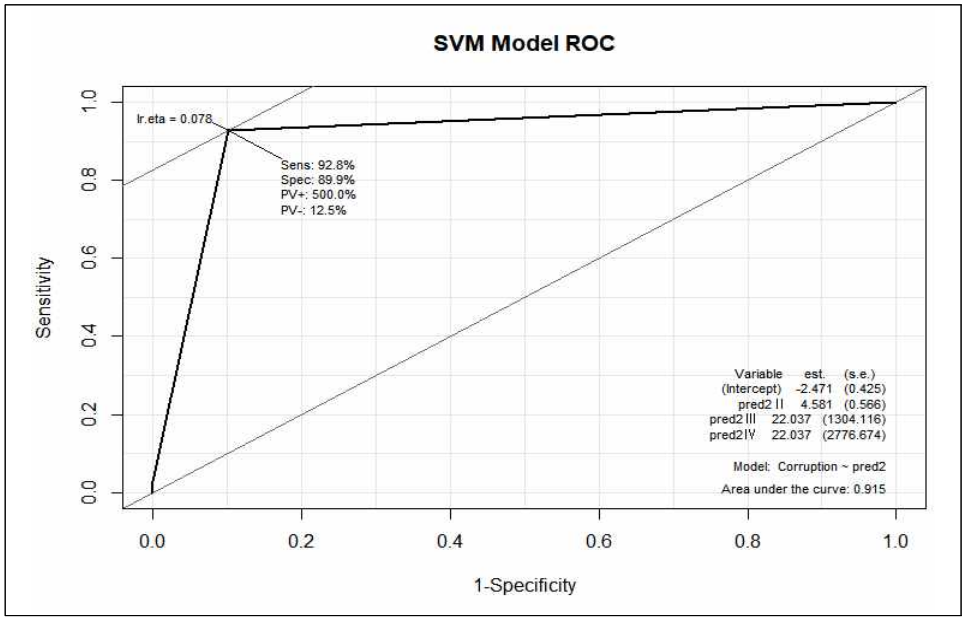
모델의 성능검증을 위해 Confusion Matrix와 ROC를 이용한 AUC를 산정하였다. 서포트 벡터 머신을 이용한 예측 모델의 Accuracy는 83.33%, AUC 91.5%로 랜덤포레스트 모델에 비해서는 성능이 떨어지는 것을 확인하였다. 특히, Ⅳ등급의 분류성능이 38%로 매우 낮은 수준이고, 모든 Class에서 Sensitivity가 낮아졌으며, 본 연구의 Multi-Class 분류모델에서는 서포트 벡터 머신을 활용하는 것이 상대적으로 부적합한 것으로 결정하였다.

```
Confusion Matrix and Statistics

          Reference
Prediction I Ⅱ Ⅲ Ⅳ
        I 71  5  1  0
        Ⅱ  8 62  4  0
        Ⅲ  0  1 54 13
        Ⅳ  0  1  6  8

Overall Statistics

               Accuracy : 0.8333
                 95% CI : (0.7793, 0.8787)
    No Information Rate : 0.3376
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.765

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: I Class: Ⅱ Class: Ⅲ Class: Ⅳ
Sensitivity            0.8987    0.8986    0.8308   0.38095
Specificity            0.9613    0.9273    0.9172   0.96714
Pos Pred Value         0.9221    0.8378    0.7941   0.53333
Neg Pred Value         0.9490    0.9562    0.9337   0.94064
Prevalence             0.3376    0.2949    0.2778   0.08974
Detection Rate         0.3034    0.2650    0.2308   0.03419
Detection Prevalence   0.3291    0.3162    0.2906   0.06410
Balanced Accuracy      0.9300    0.9129    0.8740   0.67404
```

[그림 42] Confusion Matrix(SVM_직접평가 등급)



[그림 43] Confusion Matrix2(SVM_직접평가 등급)



[그림44] ROC (SVM_직접평가 등급)

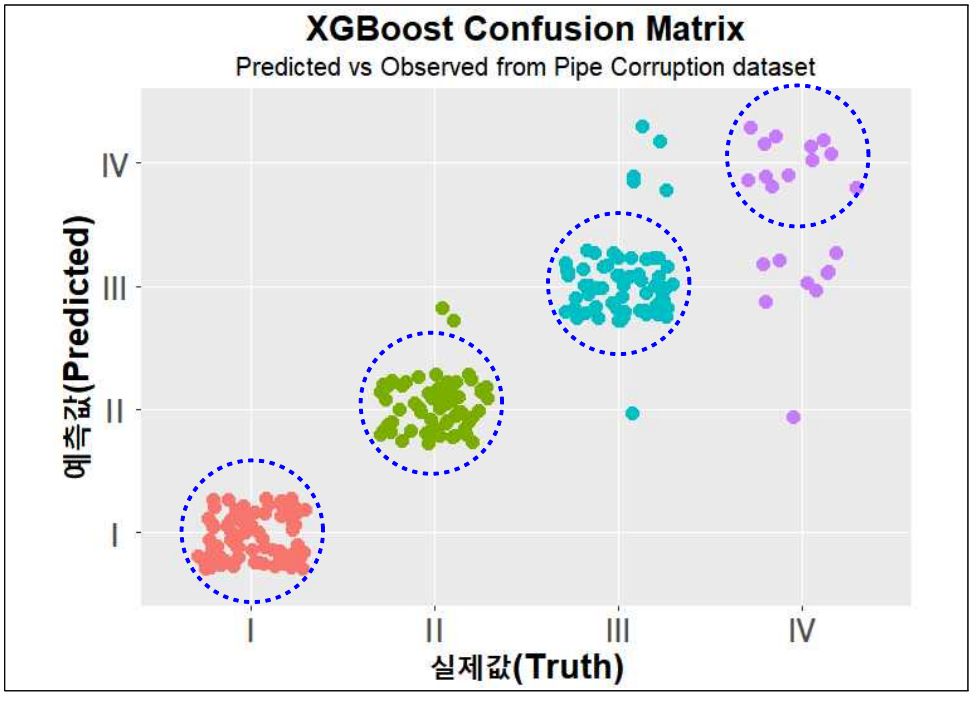
4.5.3 XGBoost를 이용한 이용한 직접평가 등급 예측

XGBoost를 이용하여 직접평가 등급을 예측하였다.

모델의 성능검증을 위해 Confusion Matrix와 ROC를 이용한 AUC를 산정하였다. XGBoost를 이용한 예측 모델의 Accuracy는 92.74%, AUC 93.48%로 랜덤포레스트 모델과 유사한 수준의 성능을 확인할 수 있다. XGBoost를 이용할 경우 Ⅳ등급의 Sensitivity가 랜덤포레스트에 비해 약 8.6% 향상된 70.6%로 개선된 것을 확인할 수 있다.

```
Confusion Matrix and Statistics

          Reference
Prediction I II III IV
        I 79  0   0  0
        II  0 67   2  0
        III 0  1  59  5
        IV  0  1   8 12

Overall Statistics

               Accuracy : 0.9274
                 95% CI : (0.8862, 0.9571)
    No Information Rate : 0.3376
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.8978

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: I Class: II Class: III Class: IV
Sensitivity            1.0000    0.9710    0.8551   0.70588
Specificity            1.0000    0.9879    0.9636   0.95853
Pos Pred Value         1.0000    0.9710    0.9077   0.57143
Neg Pred Value         1.0000    0.9879    0.9408   0.97653
Prevalence             0.3376    0.2949    0.2949   0.07265
Detection Rate         0.3376    0.2863    0.2521   0.05128
Detection Prevalence   0.3376    0.2949    0.2778   0.08974
Balanced Accuracy      1.0000    0.9794    0.9094   0.83220
```

[그림 45] Confusion Matrix(XGBoost_직접평가 등급)



[그림 46] Confusion Matrix2(XGBoost_직접평가 등급)

4.5.3 직접평가 등급 예측 결과

3가지 알고리즘을 이용하여 직접평가 등급을 예측한 결과, 랜덤포레스트와 XGBoost 모델에서 92% 수준의 정확도를 확인할 수 있었다. SVM은 Multi-Class 분류모델에서는 다른 모델에 비해 분류 성능이 떨어지므로 분석 알고리즘에서 제외하는 것으로 결정한다.

[표 20] 분류 알고리즘별 직접평가 등급 예측 성능

구 분		랜덤포레스트	SVM	XGBoost
모델의 성능	Accuracy	92.83%	83.33%	92.74%
	AUC	97.54%	91.50%	93.48%
	비고	채 택	제 외	채 택

각 Class별로 살펴보면 Class Ⅰ과 Class Ⅱ는 4.3의 노후관 개량 사업 대상여부에서 1차적으로 분류가 가능하였으며, 2차적으로 Ⅰ등급과 Ⅱ등급을 구분하기 위해 분석한 결과 높은 정확도를 갖고 있기 때문에 실무에 적용함에 있어서도 정확한 분류가 가능할 것으로 판단된다.

[표 21] 2가지 알고리즘의 Class별 Sensitivity

Sensitivity	Class Ⅰ	Class Ⅱ	Class Ⅲ	Class Ⅳ
랜덤포레스트	100%	98.57%	88.06%	61.90%
XGBoost	100%	97.10%	85.51%	70.59%

본 절까지 머신러닝을 이용한 분석을 수행하면서 독립변수의 필요성과 특성, 중요도를 분석하여 독립변수를 조정하였고, 분류 분석 모델에서 대표적으로 사용되는 3가지 알고리즘을 이용하여 각 알고리즘별 분석을 실시하였다. 변수조정과 알고리즘별 분석을 통해 Class Ⅳ의 Sensitivity가 최초 50%에서 XGBoost 모델에서 70% 수준까지 향상되었다.

하지만, Class Ⅲ은 약 85% 수준의 정확도로 비교적 양호하나, Class Ⅳ는 61~70% 수준으로 실무에 적용함에 있어 부족하다고 판단된다. 이에 다음 절에서는 직접평가 수행시 사용되는 변수로 독립변수를 조정하고, Ⅲ등급과 Ⅳ등급만을 대상으로 이진분류를 실시하여 최종 예측모델의 성능을 향상시키고자 한다.

4.6 Class Ⅲ과 Class Ⅳ의 이진분류 모델

Class Ⅲ과 Class Ⅳ의 분류 성능향상을 위해 2가지 Class에 대한 이진분류를 실시하였다. 4.4에서 1차적으로 노후관 개량사업 대상여부 예측을 통해 Class Ⅰ, Class Ⅱ그룹과 Class Ⅲ, Class Ⅳ 그룹을 분류하였기 때문에 그 결과를 바탕으로 Class Ⅲ과 Class Ⅳ의 등급을 이진분류 모델을 통해 분석하였다.

4.6.1 독립변수 재검토

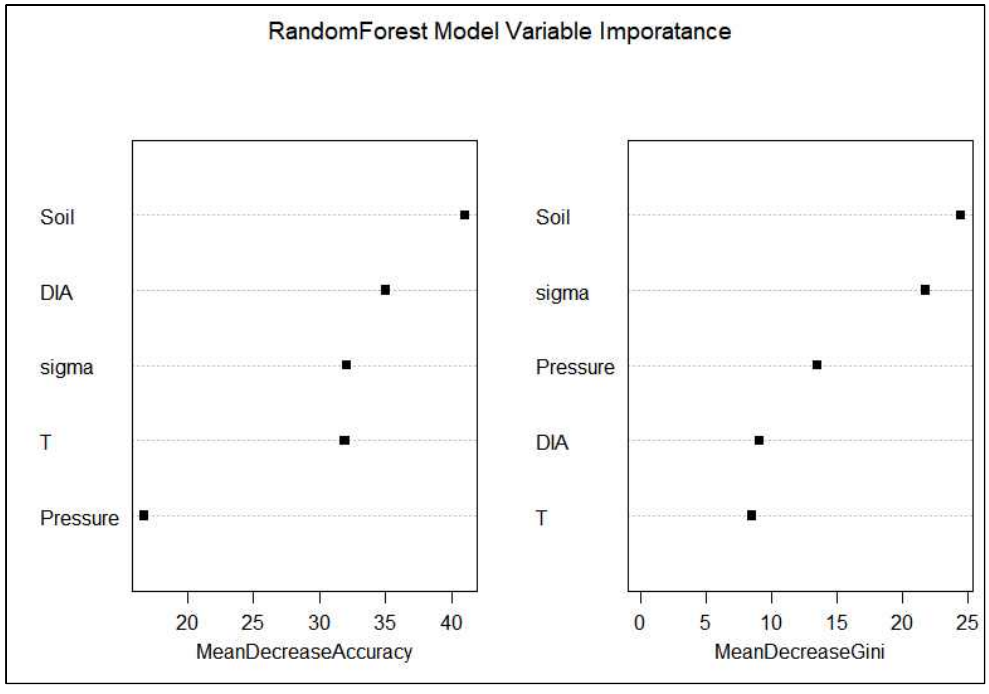
Class Ⅲ과 Class Ⅳ는 강관의 부식이 진행되어 관의 두께가 감소하고, 응력이 증가하여 안전계수가 감소한 것을 기준으로 분류하기 때문에 독립변수는 안전계수 산정시 주요 인자로 최적화하였다.

그 결과, 토양부식성, 외부 휨응력, 내부하중, 관두께, 관경 등 5가지 독립변수를 선정하였다.

[표 22] 독립변수 재검토 결과 및 변수 채택사유

간접평가 항목	독립변수 채택	채택 사유	비고
토양부식성	○	관로 부식에 영향인자로서 중요도가 지니감소 영향이 큰 인자	외부부식
외부 휨응력	○	매설 깊이와 도로의 형태를 바탕으로 외부 하중으로 인한 응력 산정값	
내부하중	○	내부 압력으로 인한 원주방향 응력 계산의 주요 인자	내부부식
관 두께	○	내부 압력으로 인한 원주방향 응력 계산의 주요 인자	
관경	○	내부 압력으로 인한 원주방향 응력 계산의 주요 인자	

토양부식성은 변수의 중요도가 가장 높은 인자로서 Mean Decrease Accuracy와 Mean Decrease Gini에서 각각 41.01, 24.46으로 변수 중에서 가장 변별력이 높은 인자이다.



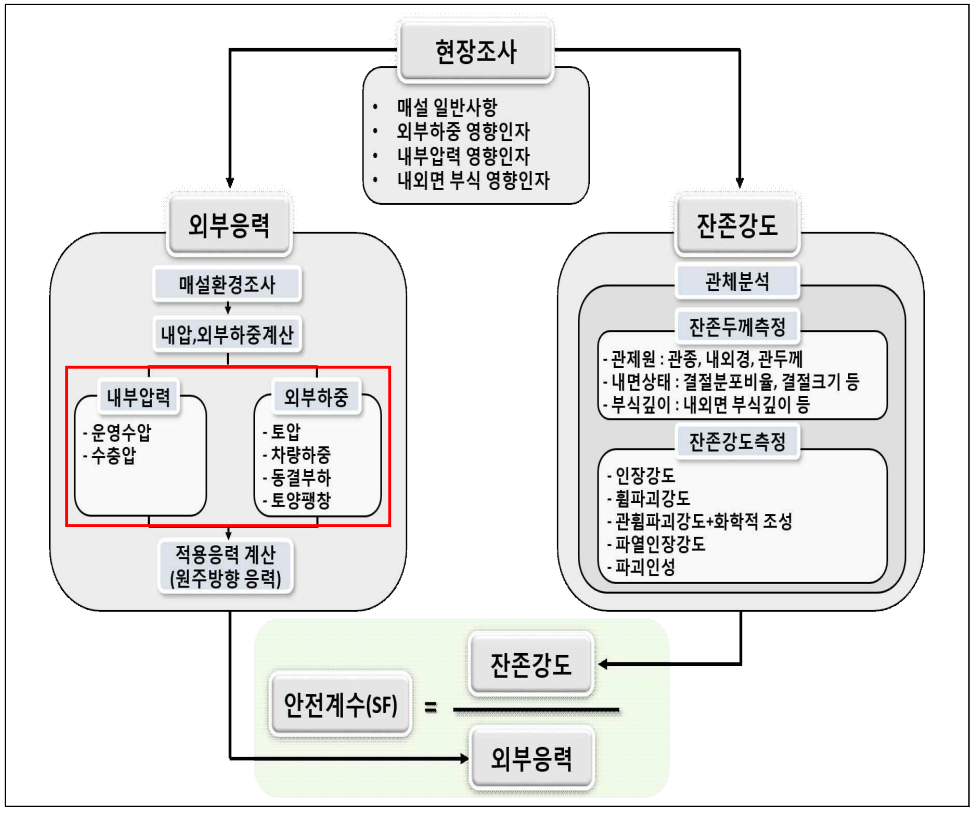
	Ⅲ	Ⅳ	MeanDecreaseAccuracy	MeanDecreaseGini
DIA	29.73273	23.53512	35.04468	9.103519
Soil	29.39208	32.44879	41.01058	24.462698
sigma	40.80750	-10.50080	32.06672	21.793688
T	27.86340	21.35430	31.93501	8.499837
Pressure	11.64813	12.60189	16.73383	13.530213

[그림 47] 독립변수 중요도 (Ⅲ, Ⅳ등급 이진분류)

외부 휨응력은 그림 48에 따라 토압(사하중)과 차량하중에 의해 계산할 수 있는 값으로서 토압은 간접평가 항목 중 매설깊이를 통해 계산할 수 있으며, 차량하중은 간접평가 항목 중 도로형태와 매설깊이에 따라 활하중으로 계산하여 반영할 수 있다.

내부하중, 관두께(초기), 관경은 내부압력으로 인한 원주방향 응력 계산의 주요인자이다. 내부수압에 의해서 원주방향으로 응력이 발생하게 되면 강관은 원주 방향의 인장응력에 대하여 견딜 수 있는 최소 관 두께를 가지고 있어야 한다. 관로의 노후화로 인하여 관의 부식이

진행되어 관 두께가 얇아지게 되면 원주방향의 응력에 견딜 수 없게 되어 파손에 이르게 된다. 따라서, 위 3가지 항목은 내부 하중에 의한 원주방향 응력의 계산공식에 사용되는 주요 인자이므로 독립변수로 채택하였다.



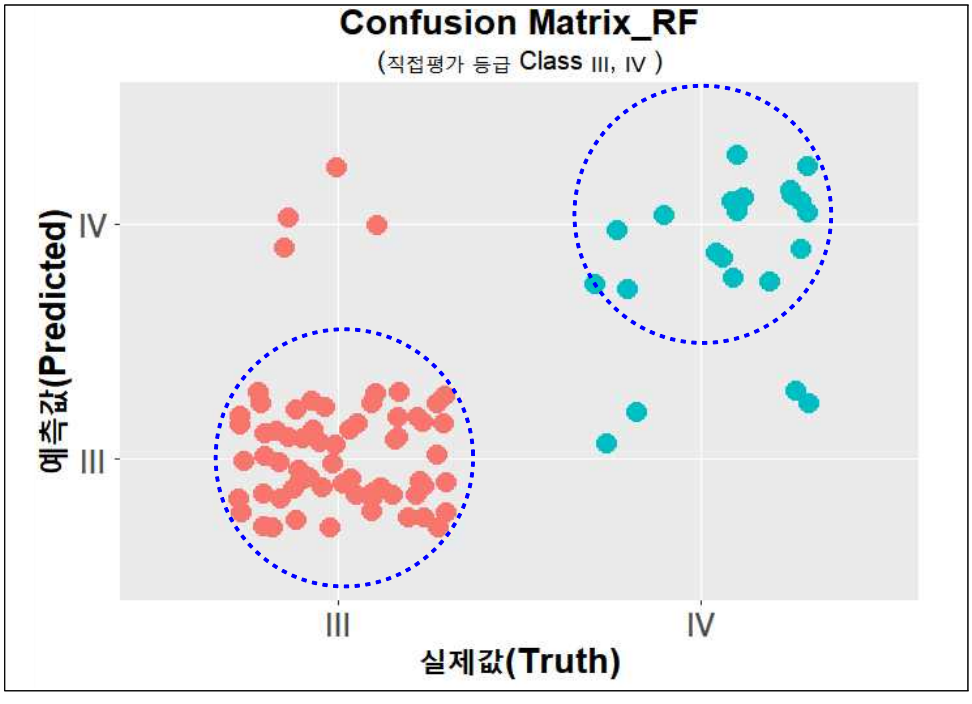
[그림 48] 강관의 안전계수 계산 흐름도[13]

4.6.2 랜덤포레스트를 이용한 Class Ⅲ과 Class Ⅳ 이진 분류

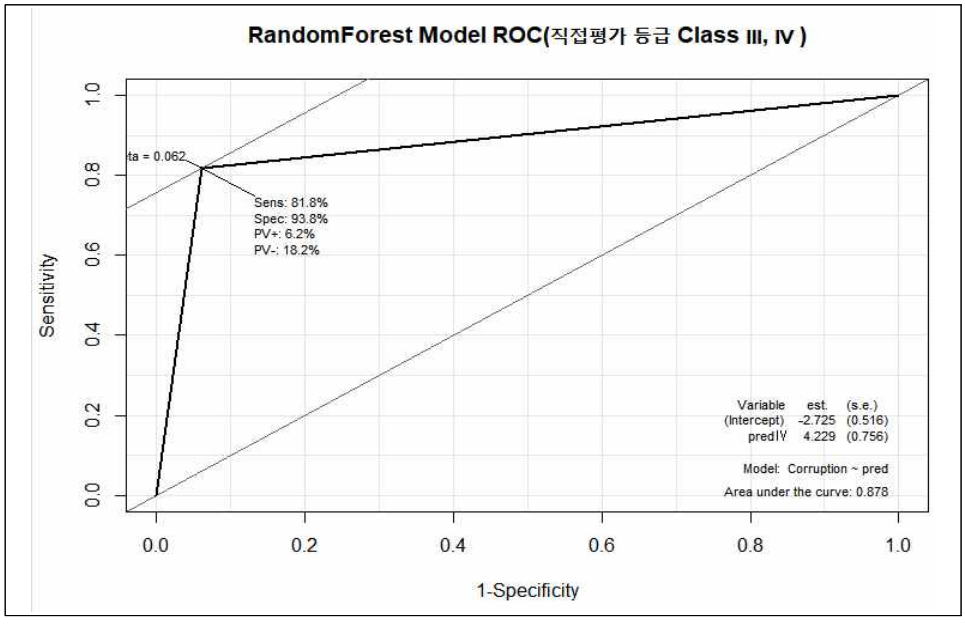
랜덤포레스트를 이용하여 Class Ⅲ과 Class Ⅳ 등급을 예측하였다. 그 결과, Accuracy는 90.80%, AUC 87.8%로 향상되었다. 특히, Class Ⅳ 등급 분류(Neg Pred Value)가 81.82%로 향상되었다. 독립변수를 직접평가 안전계수 산정시 영향을 미치는 인자로 조정하고, 이진분류로 변경한 결과 Multi-Class 분류의 61.9% 대비 약 20% 이상 향상된 성능을 확인할 수 있다.

```
Confusion Matrix and Statistics

          Reference
Prediction Ⅲ Ⅳ
        Ⅲ 61  4
        Ⅳ  4 18

               Accuracy : 0.908
                 95% CI : (0.8268, 0.9595)
    No Information Rate : 0.7471
    P-Value [Acc > NIR] : 0.0001353

                  Kappa : 0.7566

 Mcnemar's Test P-Value : 1.0000000

            Sensitivity : 0.9385
            Specificity : 0.8182
         Pos Pred Value : 0.9385
         Neg Pred Value : 0.8182
             Prevalence : 0.7471
         Detection Rate : 0.7011
   Detection Prevalence : 0.7471
      Balanced Accuracy : 0.8783

       'Positive' Class : Ⅲ
```

[그림 49] Confusion Matrix(RF_직접평가 Ⅲ, Ⅳ등급)



[그림 50] Confusion Matrix2(RF_직접평가 Ⅲ, Ⅳ등급)



[그림 51] ROC (RF_직접평가 Ⅲ, Ⅳ등급)

4.6.3 XGBoost를 이용한 Class Ⅲ과 Class Ⅳ 이진 분류

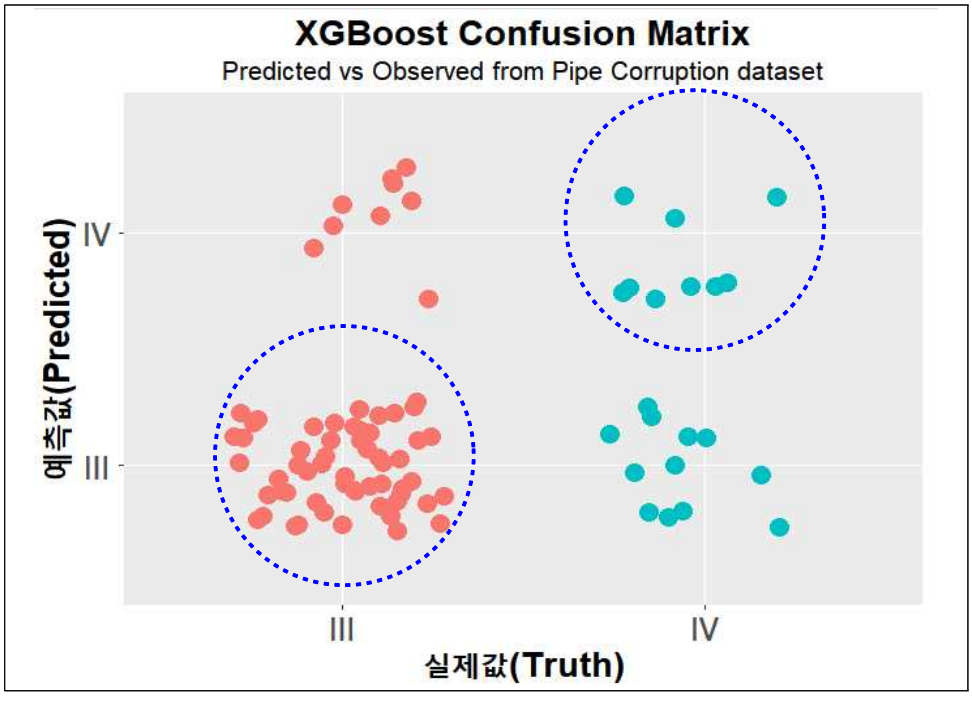
XGBoost를 이용하여 Class Ⅲ과 Class Ⅳ 등급을 예측하였다.

그 결과, Accuracy는 75.58%, AUC 64.5%로 랜덤포레스트 모델에 비해 정확도와 AUC 모두 낮았다. Class Ⅳ 등급 분류(Neg Pred Value)가 Multi-Class 분류의 57.1% 대비 약 15% 이상 하락하였다.

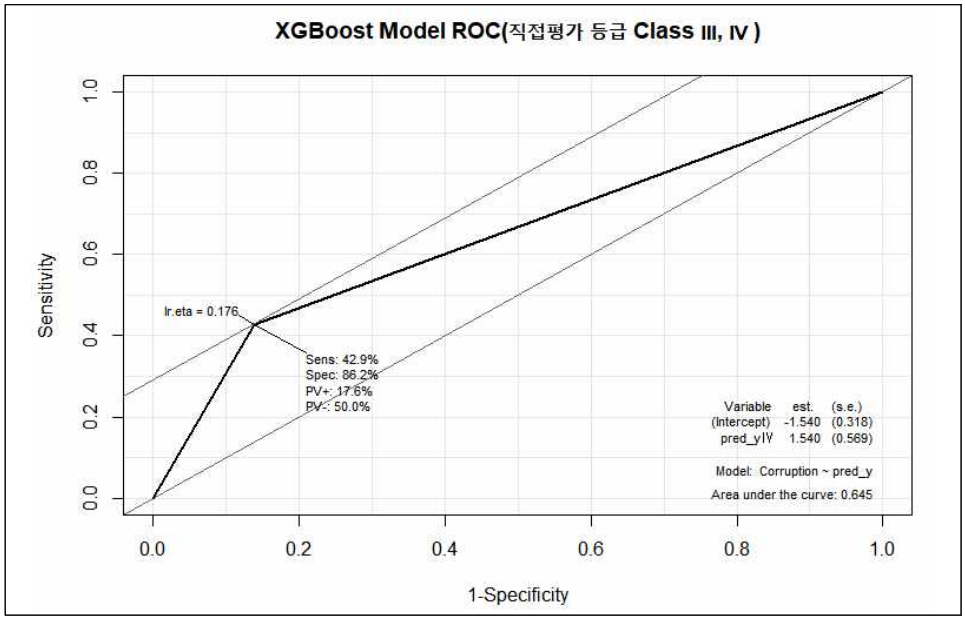
특히, Ⅳ등급으로 예측하였으나, 실제는 Ⅲ등급인 비율이 57%로 이는 사업비 산정시 과다추정의 원인이 될 수 있다. 반대로 Ⅲ등급으로 예측하였으나, 실제는 Ⅳ등급인 비율도 50%로 높아 전반적인 모델의 성능이 실무에 적용하기에는 적합하지 않는다고 판단된다.

```
Confusion Matrix and Statistics

          Reference
Prediction Ⅲ Ⅳ
        Ⅲ 56  9
        Ⅳ 12  9

               Accuracy : 0.7558
                 95% CI : (0.6513, 0.842)
    No Information Rate : 0.7907
    P-Value [Acc > NIR] : 0.8243

                  Kappa : 0.3048

 Mcnemar's Test P-Value : 0.6625

            Sensitivity : 0.8235
            Specificity : 0.5000
         Pos Pred Value : 0.8615
         Neg Pred Value : 0.4286
             Prevalence : 0.7907
         Detection Rate : 0.6512
   Detection Prevalence : 0.7558
      Balanced Accuracy : 0.6618

       'Positive' Class : Ⅲ
```

[그림 52] Confusion Matrix(XGBoost_직접평가 Ⅲ, Ⅳ등급)



[그림 53] Confusion Matrix2(XGBoost_직접평가 Ⅲ, Ⅳ등급)



[그림 54] ROC (XGBoost_직접평가 Ⅲ, Ⅳ등급)

4.6.4 Class Ⅲ과 Class Ⅳ 이진 분류 결과

Class Ⅲ과 Class Ⅳ 등급을 2가지 알고리즘으로 예측한 결과, 랜덤포레스트 모델이 XGBoost 모델에 비해 적합한 것으로 검증하였다.

이에, 강관의 노후도 예측에 있어서 랜덤포레스트 모델은 부트스트랩핑 방법으로 과적합을 방지하고, 정확도 높은 예측 결과를 나타냄에 따라 본 연구의 알고리즘은 랜덤포레스트로 결정한다.

[표 23] 2가지 알고리즘별 Ⅲ, Ⅳ등급 분류 성능

구 분		랜덤포레스트	XGBoost
모델의 성능	Accuracy	90.80%	75.58%
	AUC	87.80%	64.50%
	비고	채택	제외

랜덤포레스트를 이용한 이진 분류 결과, Accuracy가 90.80%로 매우 향상되었으며, AUC도 87.80%로 확인되었다. 특히, Class Ⅳ의 분류성능이 81.8% 수준까지 향상되어 최초 연구 단계의 50% 수준의 분류 정확도 대비 상당한 수준의 정확도 향상을 기할 수 있었다. 다만, Class Ⅳ는 분석 데이터의 수가 적어 한계점을 갖고 있으므로 지속적인 데이터 수집을 통해 분석의 정확도 향상을 기하여야 한다.

4.7에서는 본 연구에서 개발한 노후도 예측 모델을 실제 진행중인 사업의 노후도 조사에 적용하여 모델의 실무 적용가능성을 검증하고자 한다.

4.7 실제 노후관 개량사업 적용 결과

연구모델 성능을 검증하고, 실제 노후관 개량사업에 적용 가능성을 살펴보기 위해 진행 중인 노후관 개량사업의 실시설계에 본 연구 결과를 적용하였다.

진행중인 2개 노후관 개량사업은 실시설계를 진행중에 있다. 노후도 평가를 위해 간접평가와 직접조사 결과 146개 데이터를 수집하였다. 전체 221개 데이터 중 간접평가 1등급 구간, 평가 제외구간, 다른 관종인 구간, 인접지점의 직접조사를 수행한 경우는 제외하였다. 금회 개발한 연구모델에 위 데이터를 입력하여 직접평가 등급을 예측하였다.

그 결과, 표 23과 같이 직접평가 등급은 98.6%의 정확도가 있는 것으로 확인하였다. 하지만, 실제 Ⅳ등급으로 판정된 구간이 전체 146건 중 2%인 3건으로 매우 적어 실제 Ⅲ등급과 Ⅳ등급을 분류함에 있어 어느 정도 성능을 발휘하는지는 금회 검증이 곤란하였다. 다만, Ⅳ등급으로 예측한 2개 구간 모두 실제값과 일치하였으며, 나머지 Ⅰ~Ⅲ등급은 매우 높은 정확도가 있어 Ⅳ등급 데이터가 축적된다면 실무에 매우 효과적으로 이용될 수 있을 것이라고 판단된다.

[표 24] 실제 노후관 개량사업 적용 결과 오분류표

실제값 / 예측값	Ⅰ등급	Ⅱ등급	Ⅲ등급	Ⅳ등급
Ⅰ등급	42	-	-	-
Ⅱ등급	-	6	1	-
Ⅲ등급	-	-	94	1
Ⅳ등급	-	-	-	2

제5장 결론

본 장에서는 연구결과와 연구의 한계점, 연구의의와 향후 연구방향을 제시하는 등 본 연구의 결론을 제시한다. 본 연구를 통해 개발한 노후도 예측모델을 실제 노후관 개량사업에 점진적으로 확대 적용하여 당초 연구의 목적인 직접조사 없이 노후도 평가가 가능한 기반을 마련하였고, 실무에 적용 가능한 수준의 결과를 도출하였다.

5.1 연구결과

본 연구는 광역상수도 강관의 노후도 평가에 있어 기존의 간접평가와 직접평가 방식의 한계점을 극복하고, 간접평가 데이터를 이용하여 단수나 위험이 뒤따르는 직접조사 없이도 노후도 평가 등급을 예측하는 머신러닝 기반의 노후도 예측모델을 제시하였다. 그 결과, 직접조사 없이도 직접평가 등급을 92.83%의 정확도로 예측하였다.

먼저, 기존 노후도 평가방식의 한계점을 분석하고, 간접평가 항목 10가지 중 데이터의 특성을 분석하여 독립변수로서의 적정성을 판단하였다. 데이터 특성 분석을 통해 독립변수를 기존 10가지에서 5가지 항목으로 조정하고, 노후관 개량사업 대상여부와 직접평가 등급을 각각 독립변수로 하는 머신러닝 모델을 개발하였다. 또한, 분류분석 모델에서 가장 일반적으로 사용되는 랜덤포레스트, 서포트 벡터 머신(SVM), XGBoost 기법을 적용하여 본 연구에서 가장 적합한 알고리즘을 찾아 랜덤포레스트 기법을 최종 적용 모델로 채택하였다.

첫 번째 종속변수인 노후관 개량대상여부 예측에서는 97.89%의 Accuracy를 갖는 모델을 개발하였으며, 두 번째 종속변수인 직접평가 등급 예측에서는 92.83%의 Accuracy를 갖는 모델을 개발하였다. 특히,

Multi Class 분류분석에서는 Class Ⅲ과 Ⅳ의 분류 정확도가 낮은 점을 개선하기 위하여 독립변수를 직접평가의 안전계수 산정에 주요 인자로 적용되는 5가지 항목으로 최적화함으로서 분류 정확도를 개선하였다.

이에, 실무에 적용 가능한 수준의 연구결과를 얻게 되었으며, 연구 성과 검증을 위해 실제 진행중인 노후관 개량사업의 실시설계에 적용하였다. 그 결과, 직접평가 등급 예측은 98.6% 수준의 정확도가 있는 것을 확인하였다. 다만, Ⅳ등급 판정 건수가 매우 적어 Ⅲ등급과 Ⅳ등급을 분류함에 있어 어느 정도 성능을 발휘하는지는 금회 검증이 곤란하였다.

Ⅲ등급과 Ⅳ등급을 분류함에 있어 각 등급으로 판정된 건을 각 100m로 가정하고, 표9의 단위 공사비를 적용하여 정확도 개선 효과를 정량적으로 판단하였다. 정확도가 83.3%상대적으로 낮았던 SVM 모델은 Ⅲ등급 추정에서 100% 정확도가 있을때에 비해 약 16.3% 과다하게 사업비를 예측하는 반면, 정확도가 92.83%인 랜덤포레스트 모델은 7.7% 과다 추정되어 사업비 과다 추정을 8.6% 정도 개선하는 것으로 확인되었다. Ⅳ등급 추정에서 SVM 모델은 100% 정확도가 있을때에 비해 약 24.8% 과소하게 사업비를 예측하는 반면, 랜덤포레스트 모델은 12.7% 과소 추정되어 사업비 과소 추정을 12.1% 정도 개선하는 것으로 확인되었다. 이와 같이 모델의 정확도 개선은 사업비 추정에 있어 실제 사업비에 근접한 결과값을 얻는데 기여한다는 것을 확인할 수 있었다.

[표 25] 정확도 개선에 따른 정량적 기대효과

구 분	Ⅲ등급(단위 : 백만원)			Ⅳ등급(단위 : 백만원)		
	100%	모델	차이	100%	모델	차이
SVM (83.33%)	11,557	13,438	1,880 (+16.3%)	4,707	3,540	−1,166 (−24.8%)
RF (92.83%)	11,047	11,894	847 (+7.7%)	5,648	4,929	−719 (−12.7%)

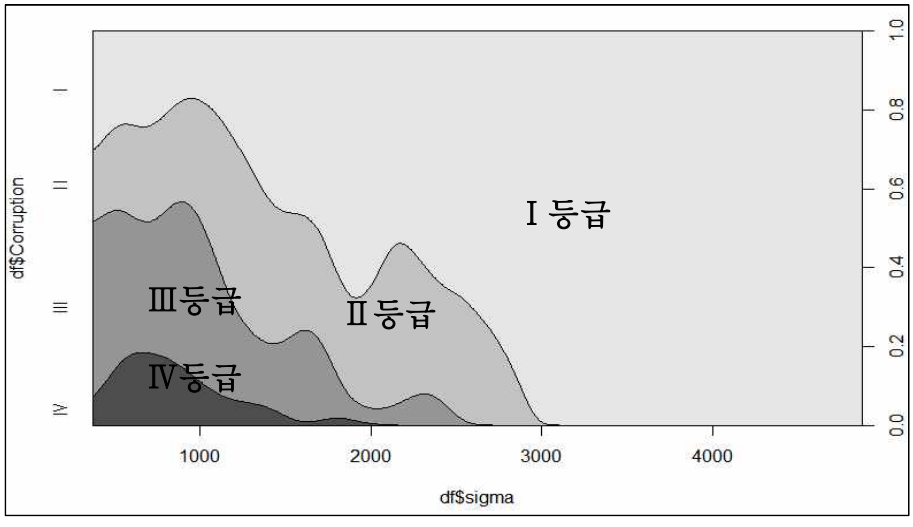
뿐 만 아니라 분석을 위한 데이터 특성을 살펴보면 향후 광역상수도 시설의 설계, 시공, 운영단계에서 고려할 특징을 찾을 수 있었다.

첫째, 관경은 노후도 평가에서 중요한 변수이며, 동일한 부식조건인 경우 관경이 작을수록 직접평가 등급이 높은 것(노후도가 큼)을 확인할 수 있었다. 평균 매설년수가 약 38년 수준인 데이터 246건을 비교해보면 평균 직경이 작은 관에서 노후도 Ⅳ등급이 많이 분포하고 있다. 사업계획 수립 시 관경은 공급유량에 따라 결정하고, 관 두께는 관경과 허용수압을 고려하여 결정한다. 하지만 향후 관의 장기 사용성 향상을 위해 노후도를 고려하면 동일한 관경에서 더 두꺼운 A종관 사용을 LCC 관점으로 검토할 수 있다.

둘째, 매설년수는 노후도에 가장 큰 영향을 미치는 인자이지만, 노후관 개량이 필요한 관로(Ⅲ등급 이상)에서는 매설년수에 의한 노후 영향을 찾기 어려웠고, 오히려 매설깊이나 토양부식성 등 다른 인자에 의해 복합적인 영향을 받는 것으로 확인되었다. 이는 관로가 경년화되면서 내외부의 피복이 손상되어 관체의 손상이 개시되면 급격하게 노후화가 진행되는 특성 때문으로 판단된다. 따라서, 최초 시공단계에서 관체의 손상이 발생하지 않도록 시공관리를 철저히 하고, 내부 용접 시 내부 도장재의 손상이 발생하지 않도록 하는 등의 품질관리가 매우 중요함을 시사한다.

셋째, 매설깊이와 관에 작용하는 외부 휨응력이 작은 경우에 노후도가 높은 것을 확인되었다. 외부 휨응력이 클 경우 노후도가 높을 것으로 예상하였으나, 실제 데이터는 그 반대의 결과임을 확인할 수 있었다. 이는 현재 외부하중에 의한 응력계산이 토압(사하중)과 차량하중(활하중)을 바탕으로 계산하여 매설깊이가 깊을수록 커지는 형태이기 때문이다. 하지만 선행연구를 살펴보면 매설깊이가 깊어 질수록 사하중은 증가하

지만 상부의 집중하중이나 등분포하중에 의한 영향은 감소하고 반대로 매설깊이가 낮을수록 사하중은 감소하지만 상부의 집중하중 등의 영향을 가장 크게 작용함을 알 수 있다. 따라서, 관로의 매설깊이 결정시 경제성 뿐 아니라 향후 관로의 노후도 증가를 고려한 적정한 매설깊이 결정이 필요함을 시사한다.



[그림 55] 외부 하중에 의한 휨응력 Density plot

넷째, 내부하중 압력이 높은 구간보다 내부압력은 낮지만 유속이 빠른 구간의 노후도가 높은 것을 확인할 수 있었다. 유속이 빠를수록 관 벽면에서의 마찰력이 증가하여 도장재 또는 관체의 부식을 가속화시키기 때문이라고 판단된다. 따라서, 광역상수도 설계시 상수도 시설기준에 따른 적정 유속 범위 내에서 시설이 운영될 수 있도록 가압시설의 양정을 고려해야 함을 시사하고 있다.

이와 같이 본 연구는 머신러닝 모델을 활용하여 노후도 평가 등급을 직접 예측하여 효과성을 제고하였고, 관로의 노후화에 영향을 미치는 변수의 특성을 살펴 설계, 운영, 시공단계에서 주의할 사항을 도출할 수 있었다.

5.2 연구의 한계점

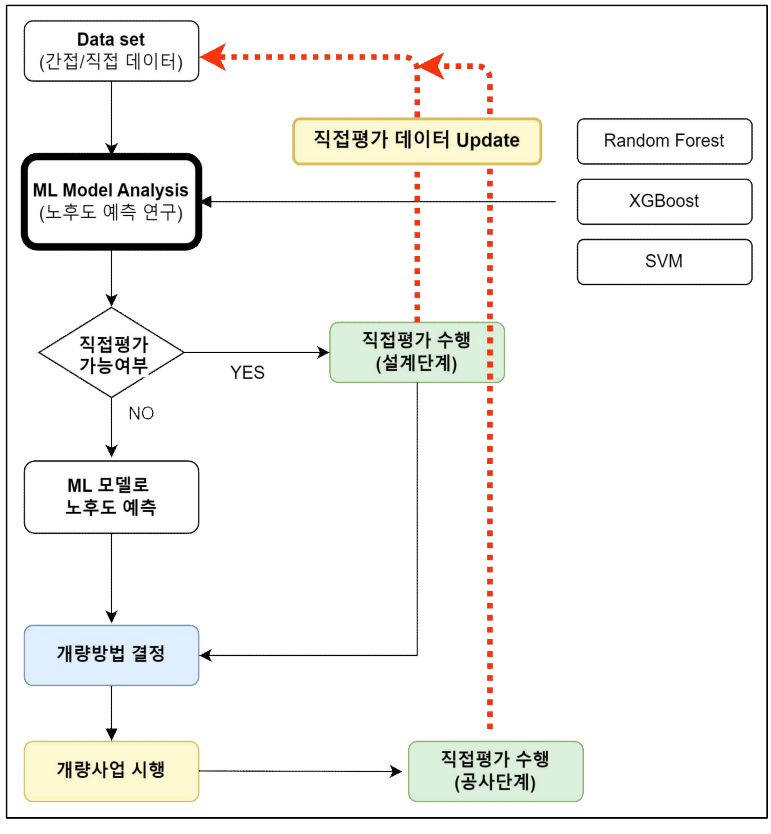
본 연구는 그간 진행된 노후관 개량사업의 데이터를 수집하여 머신러닝으로 분석하였다. 간접평가는 운영 데이터를 활용하기 때문에 데이터의 취득과 수집이 용이한 반면, 종속변수로 사용되는 직접평가 결과 데이터는 단수가 수반되고 위험이 뒤따르는 문제로 인해 수집이 매우 어렵다. 이에 따라 금회 연구에서는 787개*11항목, 8,657개의 데이터를 사용하였다.

하지만, 머신러닝을 통해 학습하기에는 데이터의 수가 비교적 적고, 특히 Class 별로 보면 Ⅳ등급 데이터는 72개로 전체 데이터의 9.2%로 매우 적었다. 본 연구의 궁극적인 목적은 금회 개발 모델을 실제 노후관 개량사업에 전면 적용하여 직접조사에 소요되는 예산을 획기적으로 절감하고, 조사과정의 위험과 단수 리스크를 제거하는데 있다. 이를 위해서는 개발 모델을 전국 어느 광역 및 공업용수도에 적용하더라도 일관된 결과가 도출되어야 한다. 그러나, 데이터의 수가 적은 점은 연구 성과를 후속 사업에 전면적으로 적용하기에는 아직 한계점이 많다.

이에, 향후 노후관 개량사업을 시행하면서 실제 단수 후 관 내부에 진입하여 개량사업을 할 때 노후도 조사를 병행하도록 내부 규정을 개선하여 노후관 데이터 수집을 더욱 활성화하고자 한다. 노후관 개량공사가 착공되면 대체관로를 설치하여 물을 공급하면서 노후관 개량이 가능하기 때문에 단수의 걱정없이 조사가 가능하므로 공사단계에서 주요 지점별 노후도 조사를 실시하도록 규정을 개정할 예정이다.

이렇게 설계단계와 공사단계에서 취득한 데이터는 학습데이터로 환류하여 분석에 활용함으로서 모델의 정확도를 향상하는데 활용하는

순환형 데이터 관리체계를 구축하도록 노력할 예정이다. 다만, 모델의 정확도 확보를 위해 직접평가 데이터가 충분히 축적될 때까지는 직접조사 수행이 필요하므로 현행대비 직접조사 비율 70%, 모델활용비율 30% 수준으로 실무에 적용하고, 데이터 축적 이후 모델 활용비율을 점차 확대할 계획이다.



[그림 56] 향후 데이터 수집 및 분석 체계

5.3 연구성과

K-water는 노후관 개량 사업을 포함한 수도시설 안정화 전략계획을 수립하였고, 환경부는 2040 국가수도기본계획에 반영하여 2040년까지 약 12조 3천억 원의 막대한 재원을 투입하여 단계적으로 개량할 계획이다. 또한, 안전정책의 강화는 안전경영을 촉구하고 있다. .

이제 노후화된 시설을 적기에 개량하여 현대화하고, 사업시행과정에서 예기치 못한 안전사고를 예방해야 한다.

본 연구에서 개발한 모델을 적용하면 노후관 개량사업의 효과성을 혁신적으로 제고할 수 있다. 향후 막대한 규모의 노후관 개량사업을 시행함에 있어 단수 리스크와 조사로 인한 조사자의 안전사고 발생위험을 획기적으로 줄일 수 있다. 본연구는 데이터 분석 기반으로 최소한의 현장조사만을 수행하기 때문에 위험상황을 제거하는데 큰 기여를 할 수 있을 것으로 기대한다.

또한, 조사 과정에서 많은 예산과 전문인력 소요문제, 조사 시간 소요로 인한 사업의 효과성 저하 문제를 해결할 수 있다. 본 연구 모델을 단계적으로 적용하면 향후 광역 및 공업용수도 노후관 개량사업에서 2040년까지 67억원 조사비 예산을 줄일 수 있다. 조사에 소요되는 기간도 획기적으로 절감하여 사업완료를 단축시킴으로서 노후관을 통해 수돗물을 공급받는 국민들에게 좀 더 빠르게 맑은 물을 공급할 수 있다. 이를 통해 국민의 수돗물에 대한 신뢰도 제고에 기여할 수 있을 것으로 기대한다.

[표 26] 예산절감액 산출 근거

노후관 개량사업비 12.3조원 × 설계비 3.5% × 설계비 중 직접조사 평균비율 9.7% × 노후관 개량대상 중 강관의 비율 53.3% × 모델활용비율 30%

* 모델활용비율은 데이터 축적 이후 점차 확대할 예정임

5.4 향후 연구

본 연구는 간접평가 데이터를 기반으로 직접평가 등급을 예측하였다. 최근에는 실제 단수를 하지 않고, 관 내부를 내시경하는 부단수 내시경 기법이 매우 발전하였고, 노후관 조사에서도 활발히 적용되고 있다. 부단수 내시경 조사시 관 내부의 동영상과 사진을 촬영하고, 분석자는 동영상과 사진을 바탕으로 관 내부의 스케일이나 녹 발생, 도장재의 박리여부, 전반적인 부식 상태 등을 종합적으로 판단하여 노후도 평가에 활용한다. 이는 분석자의 주관적인 판단에 의해 결과가 달라질 수 있다.

이에 향후 연구에서는 부단수 내시경 이미지 데이터를 판독하는 CNN 등의 방법을 적용하여 실제 단수없이도 영상이나 사진자료를 이용하여 노후도를 예측함으로서 노후도 평가 방법의 다양성을 확보하는 추가 연구가 필요하다.

또한, 금회 연구는 가장 많이 사용하고 관경이 큰 강관으로 연구범위를 한정하였으나, 주철관은 강관과 부식특성이 다르고 향후 시행될 노후관 개량사업도 많기 때문에 본 연구모델을 주철관까지 확장해야 한다. 주철관은 강관과 다르게 관의 재질이 달라 전기적인 부식의 영향을 적게 받는 반면, 취성(깨지는 성질)이 높아 사고가 발생할 경우 대규모 단수로 연결될 수 있다. 내부 라이닝이나, 접합부의 시공방법이 강관과 다르기 때문에 금회 연구와 같이 데이터의 특성을 분석하여 변수를 채택하고, 분석해야 한다.

전국 광역상수도 중 강관이 53.3%, 주철관이 40.8%이므로 주철관 노후도 예측모델이 개발되면 전국 광역상수도 94.1%를 예측할 수 있는 기반이 마련된다. 이를 위한 추가 연구가 필요하다.

참 고 문 헌

[1] 2020 상수도 통계 (환경부, 2020)

[2] 상하수도 용어집 (한국상하수도협회, 2022)

[3] 상수도 관망의 기술진단 범위 및 시행방법 등에 관한 고시 (환경부 고시 제2007-11호)

[4] 도·송수시설 기술진단 가이드북(K-water, 2010)

[5] 상수관망 정밀조사 매뉴얼(환경부, 2019)

[6] 2040 국가수도기본계획(환경부, 2022)

[7] 수도시설 운영관리 편람(K-water, 2020)

[8] KS D 3565(상수도용도복장 강관) (한국표준협회, 2022)

[9] 나노복합수지 PE3층 피복강관 카달로그(웰텍주식회사, 2022)

[10] 기존 광역상수도 대구경 강관에 대한 정밀조사 보고서 (K-water연구원)

[11] 정홍도, and 김인규. "지중 매설관 부식원인과 방지대책에 관한 연구." 대한설비공학회 학술발표대회논문집 (2012): 161-164.

[12] 한금석, et al. "수도관 부식에 대한 잔류염소 농도 영향 및 부식제어 방안." Corrosion Science and Technology 17.1 (2018): 12-19.

[13] 수도권광역상수도 2단계 신뢰성 제고사업 기본 및 실시설계 보고서(K-water, 2015)

[14] 수도시설 운영비 및 개략공사비 산정기준(환경부, 2020)

[15] 밀폐공간 질식재해예방 안전작업 가이드(고용노동부, 안전보건공단, 2022)

[16] https://www.cdc.gov/niosh/index.htm, Hierarchy of controls (NIOSH)

[17] 수도시설 안정화사업 전략계획 수립 용역 보고서(K-water, 2021)
[18] 광역상수도 관로 부식두께 분석을 통한 잔존수명 산정 연구 (기존 관체 조사 자료의 통계학적 모델을 통하여) (안홍기, 2018)
[19] 고영환, et al. "전기방식 적용여부에 따른 관로 부식상태 비교 및 경제성 검토." (2007): 390-393.
[20] 류연선, 조현만, and 김영빈. "상재하중의 재하위치 및 매설심도를 고려한 지중 매설관의 구조거동 평가." 2. 한국방재학회 논문집 17 (2017): 231-236.
[21] 토질역학 이론과 응용(김상규, 2004, 청문각)
[22] 이윤호. 랜덤포레스트 기법을 이용한 상수관로 부식에 관한 연구(2020)
[23] 김기태, et al. "AI 기반 상수도시설 개량 의사결정 모델 분석." 사물인터넷융복합논문지 Vol.8 No.5, pp 11-16, 2022
[24] 김주환, et al. "노후 상수도관 관리실태와 진단을 통한 개선." 물과 미래, Vol.52 No.9, pp 6-16, 2019.9
[25] 김대웅, et al. "Assessment of Pipe Deterioration in Water Distribution Network of Incheon City Using Principal Component Analysis and Artificial Neural Network" Crisisonomy, Vol.17 No.1, pp 103-115, 2019.9
[26] 이슬민, et al. "ANN-Clustering 기법을 이용한 상수관로 노후도 평가 및 분류" J.Korea water resour. assoc, Vol.51 No.11, pp 959-969, 2018

Abstract

A Study on Prediction of Aging Evaluation Grade of Water Supply Pipeline Using Machine-Learning

Jong Ho, LEE

Graduate School of Engineering Practice

Seoul National University

As of 2019, the water supply rate in Korea was about 99.3%, and water supply services are available in most areas of the country. The number of aging pipelines has gradually increased to 79.461km, which is 34.8% of the total pipelines, and is expected to continue increasing over time.

Aging water supply infrastructure can lead to water quality problems such as interruption of water supply due to broken supply pipes, red water due to scales and corrosion problems inside the water supply system, and the distrust of water supply system.

Therefore, It is necessary to provide stable water supply service by improving aging water supply pipelines in a timely manner.

A necessary procedure for the improvement of aging water pipeline is Deterioration evaluation. Deterioration evaluation is divided into indirect evaluation and direct evaluation. Indirect evaluation is to predict the aging state of pipelines based on operating factors and Direct evaluation is to directly measure the degree of corrosion inside and outside the pipe by cutting or excavating the pipe.

However, the current indirect and direct evaluation methods have limitations, such as data representativeness, differences in evaluation grades, risks in the investigation process, and risk of interruption of water supply.

The purpose of this study is to predict the aging evaluation grade of water supply pipeline using machine learning without direct investigation process. As a result of the study, it is possible to predict the aging evaluation grade of water supply pipeline with an accuracy of 92.83%. It is expected that the results of this study will be applied to actual projects to improve the effectiveness of the project and to solve interruption of water supply and safety problems.

keywords : Deterioration evaluation, Indirect evaluation, Direct evaluation, Machine learning, Prediction of the aging evaluation grade

Student Number : 2021-20316